THEORY IN PRACTICE

数据之美

Beautiful Data

揭秘优雅的数据解决方案背后的故事

Toby Segaran & Jeff Hammerbacher 编

祝洪凯 李妹芳 段炼 译

O'REILLY®

机械工业出版社
China Machine Press

华章科技

O'Reilly精品图书系列

数据之美

Beautiful Data:The Stories Behind Elegant Data Solutions

[美]Toby Segaran

Jeff Hammerbacher　编

祝洪凯　李妹芳　段炼　译

ISBN：978-7-111-31512-4

本书纸版由机械工业出版社于2010年出版，电子版由华章分社（北京华章图文信息有限公司）全球范围内制作与发行。

客服热线：+ 86-10-68995265

客服信箱：service@bbbvip.com

官方网址：www.bbbvip.com

新浪微博 @研发书局BBBVIP

腾讯微博 @bbb-vip

□ □

□□

□□□□

□□□□□□

□□□□

□□□□□

□□□□□□□□□□□□□

□□□□□□□□□□□□

□□□□

“□□”□□□□□□□□□□□

□□□

□4□ PNUTShell□□□□□□□

□□

□□□□

□□□□

□□□□□□□□

□□

□□

□□□□

□5□ □□□□□□□□□□□□□

□□□□□□

Facebook□□□“□□□□”

[illegible]

[illegible]“[illegible]”

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]11[illegible] [illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]12[illegible] Sense.us[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

“[illegible]”[illegible]“[illegible]”

[illegible]

[illegible]

[illegible]13[illegible] [illegible]

[illegible]

[illegible]

[illegible]

[illegible]14[illegible] [illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]15[illegible] [illegible]DNA[illegible]

[illegible]DNA[illegible]

DNA[illegible]

[illegible]

DNA[illegible]

[illegible]

[illegible]16[illegible] [illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]17[illegible] [illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]18[illegible] [illegible]

[illegible]

[illegible]

[illegible]

[illegible]

□□

□□□□□□□

□□□□□□□□□

□□□□

□□□□□□

□□□□□

□□

□□□□

□19□　□□□□□□□

□□1□□□□□□□□□□□□

□□2□□□□□□□□

□□3□□□□□□

□□4□□□□□□□□□□□□□□□□□□□□□□

□□5□□□□□□□□□□□□

□□

□□□□

□20□　□□□□

□□□□□□□□□□□□□

□□□□□□□□

□□□□

□□□□□□□

O'Reilly Media,Inc.[illegible]

[illegible]O'Reilly Media,Inc.[illegible]

O'Reilly Media,Inc.[illegible]Unix[illegible]X[illegible]Internet[illegible]

[illegible]The Whole Internet User,s Guide[illegible]Catalog[illegible]20[illegible]50[illegible]GNN[illegible]Internet[illegible]WebSite[illegible]PC[illegible]Web[illegible]O'Reilly Media,Inc.[illegible]Internet[illegible]

[illegible]O'Reilly Media,Inc.[illegible]——[illegible]O'Reilly Media,Inc.[illegible]O'Reilly Media,Inc.[illegible]O'Reilly Media,Inc.[illegible]O'Reilly Media,Inc.[illegible]——[illegible]O'Reilly Media,Inc.[illegible]O'Reilly Media,Inc.[illegible]O'Reilly Media,Inc.[illegible]

[illegible]

[illegible]Yahoo[illegible]Deep Web[illegible]Facebook[illegible]“[illegible]”[illegible]DNA[illegible]Radiohead[illegible]

[illegible]Deep Web[illegible]NLP[illegible]Radiohead[illegible]

[illegible]“[illegible]”[illegible]——[illegible]“[illegible]”[illegible]“[illegible]”[illegible]

[illegible]

[illegible]

[illegible]

2010[illegible]6[illegible]26[illegible]

[illegible]

[illegible]“[illegible]”[illegible]

[illegible]

Radiohead[illegible]DNA[illegible]“[illegible]”[illegible]

[illegible](Ceative Commons)[illegible](te Sunlight Foundation)[illegible]“[illegible]”[illegible]

[illegible]

[illegible]

[illegible]1[illegible]Nathan Yau[illegible]

[illegible]2[illegible]Jonathan Follett[illegible]Matthew Holm[illegible]Web[illegible]

[illegible]3[illegible]J.M.Hughes[illegible]

[illegible]4[illegible]PNUTShell[illegible]Brian F.Cooper[illegible]Raghu Ramakrishnan[illegible]Utkarsh Srivastava[illegible]Web[illegible]

[illegible]5[illegible]Jeff Hammerbacher[illegible]Facebook[illegible]

[illegible]6[illegible]Jason Dykes[illegible]Jo Wood[illegible]

[illegible]7[illegible]Jeff Jonas[illegible]Lisa Sokol[illegible]

[illegible]8[illegible]Jud Valeski[illegible]Web[illegible]

[illegible]9[illegible]Deep Web[illegible]Alon Halevy[illegible]Jayant Madhavan[illegible]G[illegible]“[illegible]”[illegible]Web[illegible]

[illegible]10[illegible]Radiohead[illegible]“House of Cards”[illegible]Aaron Koblin[illegible]Valdean Klump[illegible]“[illegible]”[illegible]

[illegible]11[illegible]Michal Migurski[illegible]

[illegible]12[illegible]Sense.us[illegible]Jeffrey Heer[illegible]150[illegible]

[illegible]13[illegible]Coco Krumme[illegible]

[illegible]

[illegible]14[illegible]Peter Norvig[illegible]Web[illegible]1[illegible]

[illegible]

[illegible]15[illegible]DNA[illegible]Matt Wood[illegible]Ben Blackburne[illegible]

[illegible]DNA[illegible]

[illegible]16[illegible]Jean-Claude Bradley[illegible]Rajarshi Guha[illegible]

Andrew Lang[illegible]Pierre Lindenbaum[illegible]Cameron Neylon[illegible]Antony

Williams[illegible]Egon Willighagen[illegible]“[illegible]”(cowdsourcing)[illegible]

[illegible]

[illegible]17[illegible]Brendan O'Connor[illegible]Lukas

Biewald[illegible]

[illegible]18[illegible]Hadley Wickham[illegible]Deborah

F.Swayne[illegible]David Poole[illegible]

[illegible]

[illegible]19[illegible]Andrew Gelman[illegible]Jonathan P.Kastellec[illegible]Yair

Ghitza[illegible]

[illegible]20[illegible]Toby Segaran[illegible]Web[illegible]

[illegible]

[illegible]

[illegible]

[illegible](Ialic)

[illegible]URL[illegible]Email[illegible]

[illegible](Cnstant width)

[illegible]

[illegible](Cnstant width bold)

[illegible]

[illegible](Cnstant width italic)

[illegible]

[illegible]

[illegible]O'Reilly[illegible]

[illegible]ISBN[illegible]“Beautiful Data,edited by Toby Segaran and Jeff Hammerbacher.Copyright 2009 O'Reilly Media,Inc.[illegible]978-0-596-15711-1.”[illegible]

[illegible]permissions@oreilly.com[illegible]

[illegible]

[illegible]

[illegible]

O'Reilly Media,Inc.

1005 Gravenstein Highway North

Sebastopol,CA 95472

[illegible]

[illegible]2[illegible]C[illegible]807[illegible]100035[illegible]

[illegible]

O'Reilly[illegible]

[illegible]

http://www. oreilly.com/catalog/9780596157111/

[illegible]

bookquestions@oreilly. com

[illegible]

http://www. oreilly.com

http://www. oreilly.com.cn

[illegible]1[illegible] [illegible]

NathanYau

[illegible]Web[illegible]Web[illegible]

[illegible]20[illegible]30[illegible]Mass Observation[illegible]——[illegible]1930[illegible]

[illegible]Mass Observation[1][illegible]

[illegible]GB[illegible]

[illegible]

[illegible]PEIR(Personal Environmental Impact Report)[illegible]PEIR[illegible]your.flowingdata(YFD)[illegible]Twitter[illegible]

[illegible]PEIR[illegible](U)[illegible]YFD[illegible]PEIR[illegible]YFD[illegible]PEIR[illegible]YFD[illegible]PEIR[illegible]YFD[illegible]PEIR[illegible]YFD[illegible]

[illegible](PIR)

PEIR[illegible](Cnter for Embedded Networked Sensing)[illegible](Uban Sensing group)[illegible]DietSense[illegible]Family Dynamics[illegible]Walkability[illegible][2][illegible]GPS[illegible]

PEIR[illegible]Campaignr[illegible]——[illegible]2[illegible]GPS[illegible]PEIR[illegible]——GPS[illegible]“[illegible]”[illegible]1000[illegible]1000[illegible]

[1]Mass Observation[illegible]500[illegible] http[illegible]//en.wikipedia.org/wiki/Mass_Observation[illegible]

[2]CENS[illegible]http[illegible]//urban.cens.ucla.edu/[illegible]

your.flowingdata(YFD)

PEIR[illegible]YFD[illegible]Twitter[illegible]Twitter[illegible]

[illegible]YFD[illegible]SMS[illegible]

YFD[illegible]Twitter[illegible]Twitter[illegible]SMS[illegible]

[illegible]PEIR[illegible]YFD[illegible]

[illegible]

[illegible]

[illegible]

[illegible]Twitter[illegible]YFD[illegible]Twitter[illegible]Twitter[illegible](teet)[illegible]Twitter[illegible]SMS[illegible]

[illegible]E-mail[illegible]Twitter[illegible]E-mail[illegible]SMS[illegible]YFD[illegible]Twitter[illegible]E-mail[illegible]Twitter[illegible]E-mail[illegible]Twitter[illegible]Twitter[illegible]Twitter[illegible]

E-mail E-mail

——

Twitter YFD

Twitter “

”(ae salami sandwich)

SQL “ate”

“salami sandwich”

“goodnight” “gmorning”

PEIR YFD PEIR

PEIR YFD Twitter PEIR

GPS

YFD

Twitter Twitter

Twitter Twitter SMS

YFD

Twitter “ 6 ”(ae roast chicken and

potatoes at 6□00pm)□□“□□11□□□”(godnight at 23□00□□□□□□□□□□□□□Twitter□□□□□□□□□“at hh:mm”□□□YFD□□□□□□□□□□□□□□□□□□□□□“goodnight”□□□□YFD□□□□□□□□□□□□

PEIR□□□□□□□□□□“□□”(i the moment)□□□□□□□□□□□Campaignr□□□□□□□□□□□□□□GPS□□□□20□□□□□□□□□□□□□□□□□□□□□□□□PEIR□□PEIR□□□100%□□□

□□□

[illegible]

[illegible]YFD[illegible]PEIR[illegible]YFD[illegible]MySQL[illegible]YFD[illegible]——[illegible]PEIR[illegible]

[illegible]PEIR[illegible]PostGIS[illegible]PostgreSQL[illegible]MySQL[illegible]GIS[illegible]PEIR[illegible]PostGIS[illegible]PostgreSQL[illegible]

[illegible]10[illegible]PEIR[illegible]PEIR[illegible]PEIR[illegible]PEIR[illegible]PEIR[illegible]PEIR[illegible]c onsolidation)[illegible]

□□□□□□□□□YFD□□□□□□□YFD□□Django□□□□□Python□□□Web□□□□□□□MVC(Model-View-Control□□□-□□-□□□GIS□□□□□□□□□□□PEIR□□□□□□□□□——□□

[illegible]

[illegible]YFD[illegible]PEIR[illegible]

PEIR[illegible]GPS[illegible]PEIR[illegible][1][illegible]

1.[illegible](prcel data)[illegible]

2.[illegible]Web[illegible]

3.[illegible]Web[illegible]

4.[illegible]

[illegible]PEIR[illegible]

[1]PEIR,http：//peir.cens.ucla.edu

[illegible]

[illegible]YFD[illegible]PEIR[illegible]

[illegible]“[illegible]”[illegible]

PEIR

[illegible]PEIR[illegible]——“[illegible]”[illegible]1000[illegible]PEIR[illegible]

[illegible]

□□□□□□□□□□PEIR□□□□□□□□□□□□□□□□□□□□□GMap API□□□□□□□□□□□□□□□□——□□Modest Maps□□□□Modest Maps□□□“□□”(tlebased)□□□□□□□□□□□Flash□□□□□□ActionScript 3.0□□□Modest Maps□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□Modest Maps□□□□□□□□□□□□□□“□□”□□□□“□□”□□□Modest Maps□□□□□□□□□□

□□□□□(vsual cue)□□□

□□□□□□□□□□□□□□□□□□□□□□□PEIR□□□□□□□□□□□□□□□□

·□□□□□□□□□□□□□□□□□□□□□□□□□□

·□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

·□□□□□□□□□□□□□□□□□□□

·□□□□□□□□□GPS□□□□□□□

·□□□□□□□□□□□□□□□□□□□“□□”□

□□□□□□□□

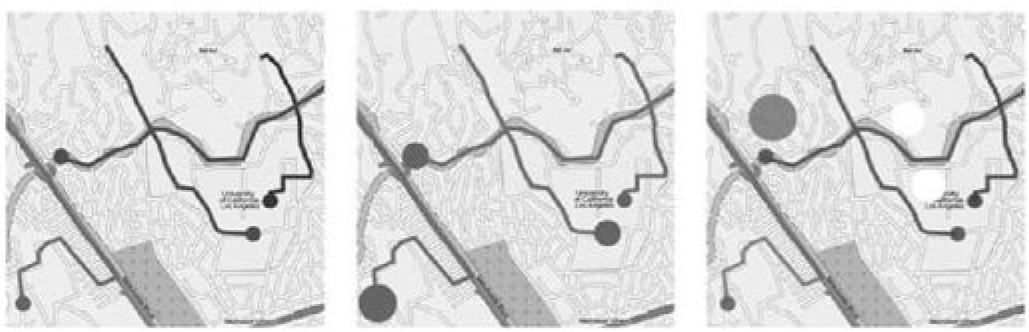
□□□□□□□□□□□GPS□□Modest Maps□OpenStreetMap□“□□”□□□□□□□□□□□GPS□□□□□□□□□□□□□□□□□□□□□□□□□(crcle)□□□□□□□□□□□□□□□□□□□□□

[illegible]——[illegible]

[illegible]

[illegible]1-1[illegible]

[illegible]

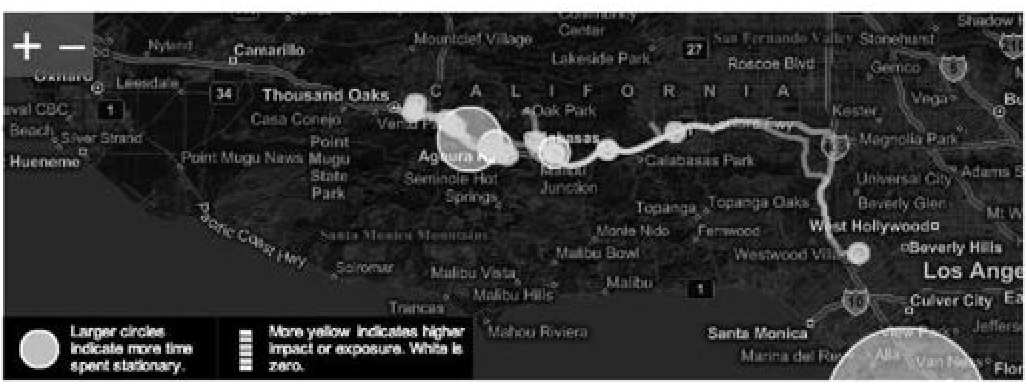


[illegible] 1-1[illegible]GPS[illegible](aea circle)[illegible]GPS[illegible]1[illegible]

[illegible]

[illegible]GPS[illegible]GPS[illegible]

[illegible]Trulia Snapshot[illegible]“[illegible]”[illegible]1-2[illegible]Modest Maps[illegible]



[illegible] 1-2[illegible]GPS[illegible]

[illegible]2[illegible]

[illegible]

[illegible]“[illegible]”[illegible]

[illegible]

[illegible]

[illegible]PEIR[illegible]5

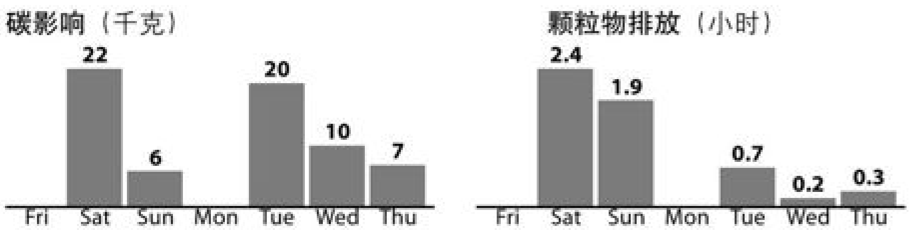
[illegible]Cabspotting[illegible]

[illegible]

[illegible]

PEIR[illegible]

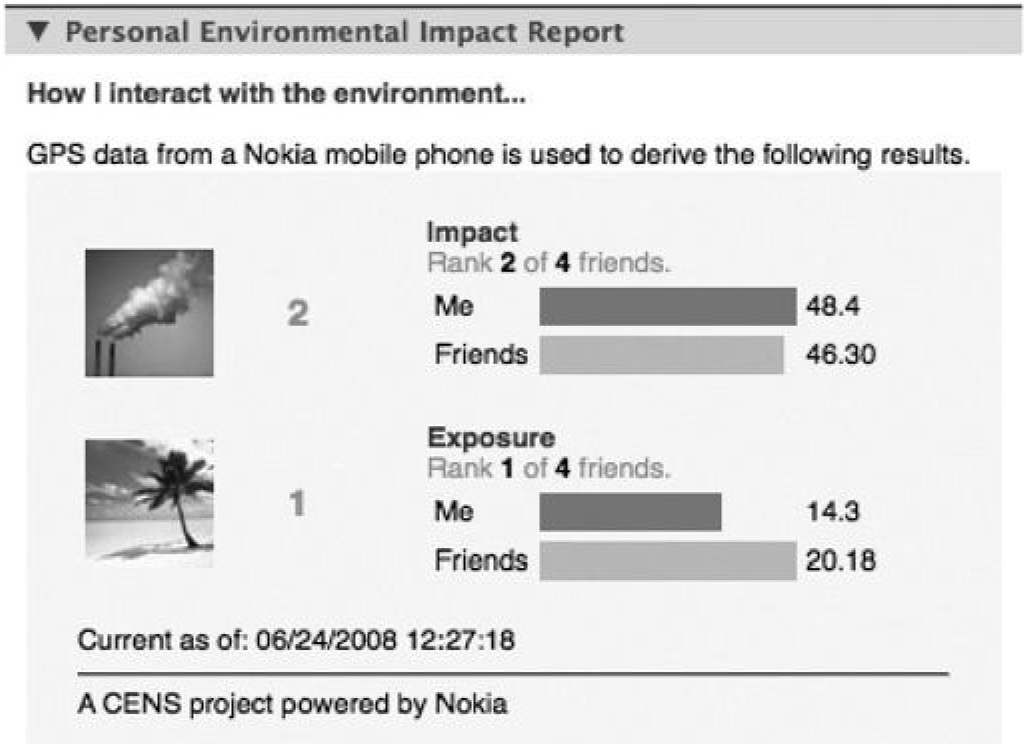
[illegible]1-3[illegible]



[illegible] 1-3[illegible]

[illegible]

PEIR[illegible]Facebook[illegible]1000[illegible]Facebook[illegible]PEIR Facebook[illegible]Facebook[illegible]1-4[illegible]



[illegible] 1-4 [illegible] PEIR [illegible] Facebook [illegible] 3 [illegible]

PEIR Facebook [illegible]

[illegible] PEIR [illegible] PEIR [illegible]

[illegible]GPS[illegible]

YFD

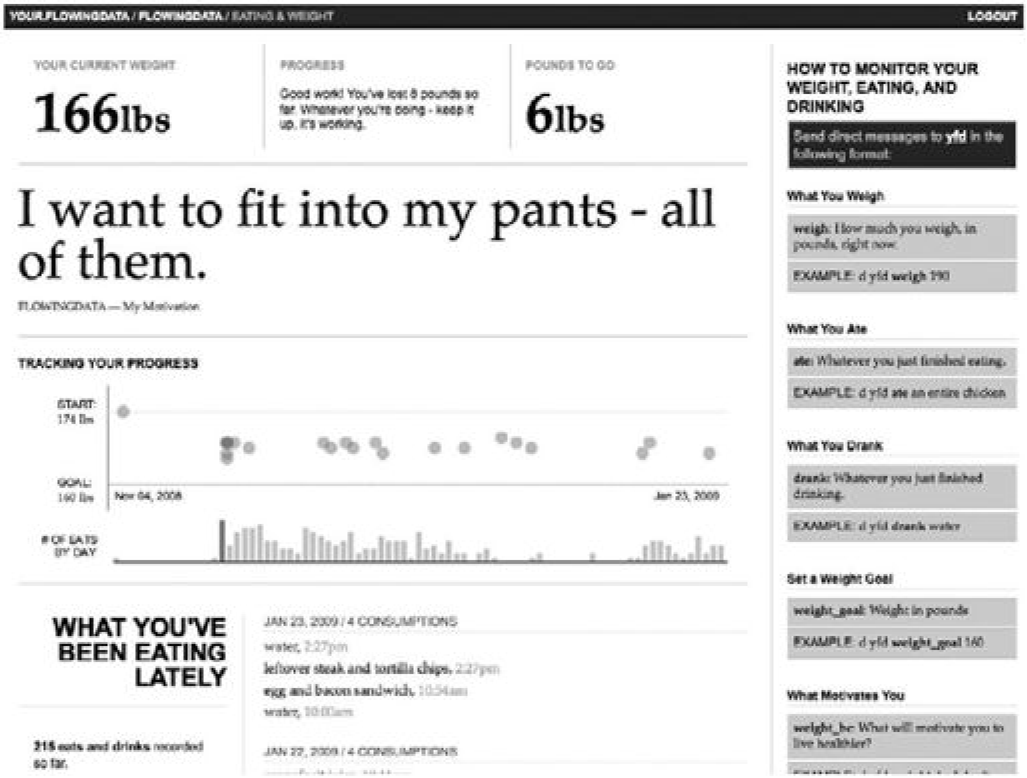
PEIR[illegible]YFD[illegible]YFD[illegible]PEIR[illegible]YFD[illegible]

[illegible]Twitter[illegible]Twitter[illegible]YFD[illegible]1-5[illegible]

[illegible]Twitter[illegible]YFD[illegible]Twitter[illegible]YFD[illegible]Twi tter[illegible]Twitter[illegible]YFD[illegible]PEIR[illegible]YFD[illegible]

[illegible](felings and emotions)[illegible]1-6[illegible](srted tag cloud)[illegible]

□□□□□□□□□□□□□□□□□□□□□YFD□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□YFD□□



□ 1-5□□□□□□□□□□□□□□□□YFD□□□□□□□□□□□□□□□4□



[illegible] 1-6[illegible]YFD[illegible]

[illegible]“[illegible]”(coud)[illegible]

[illegible]

[illegible]

□□□□

PEIR□YFD□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□
http□//peir.cens.ucla.edu□http□//your.flowingdata.com□□□□
□□□PEIR□□□□□□□□□□□□□□□□□□□□□□□□□□□□□YFD□□□□
□□□□□□□□□□□□□□

□2□　□□□□□□□□□□□□□□□□□□□

Jonathan Follett□Matthew Holm

[illegible]

[illegible](uer experience,UX)[illegible]——[illegible]“[illegible]”[illegible]800[illegible]

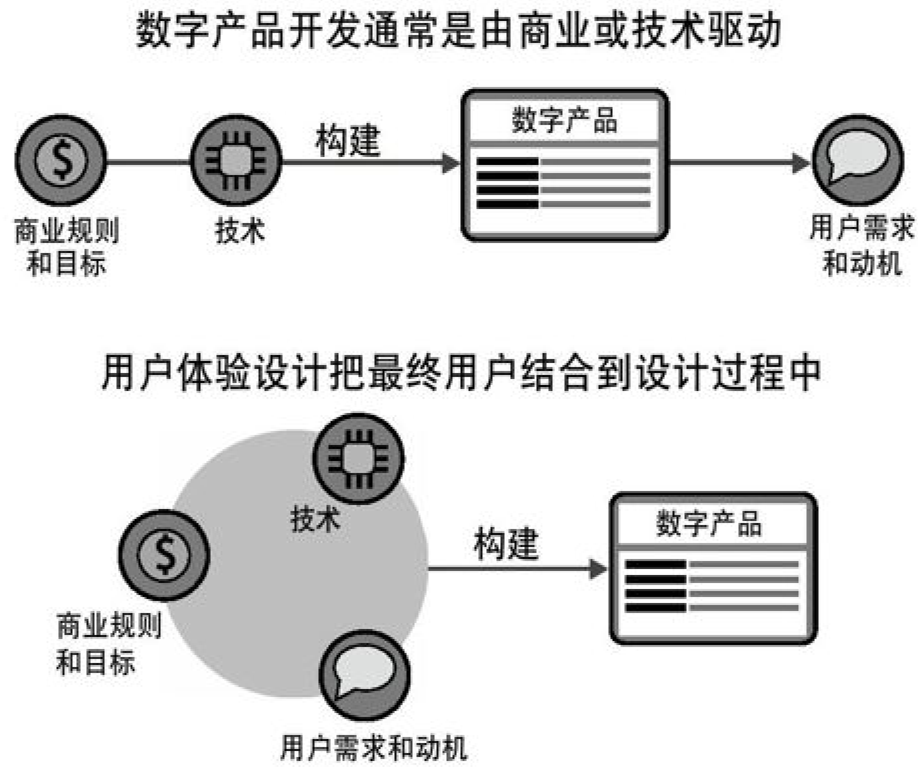
[illegible]

[illegible]2-1[illegible]Web[illegible]

[illegible]

[illegible]

[illegible]Web[illegible]


数字产品开发通常是由商业或技术驱动
商业规则
和目标
技术
构建
数字产品
用户需求
和动机
用户体验设计把最终用户结合到设计过程中
技术
商业规则
和目标
构建
数字产品
用户需求和动机

[illegible]

[illegible]Urban Wallace Associates[illegible]55[illegible]75[illegible]

[illegible]Urban Wallace Associates[illegible]“[illegible]35%[illegible]”UWA[illegible]Roger Urban[illegible]“[illegible]80%[illegible]”

[illegible]“[illegible]”(gtekeeper)[illegible]

Urban Wallace Associates[illegible]“[illegible]”(lnding page[illegible]Web[illegible]

[illegible]

[illegible](acessibility)[illegible]

[illegible]

[illegible]

[illegible]Web[illegible]72%[illegible]45[illegible]

[illegible]——[illegible]

[illegible]

[illegible]

[illegible]

Internet[illegible]

“[illegible]”

“[illegible]”

“[illegible]”

“[illegible]……[illegible]……[illegible]”

[illegible]——[illegible]Roger Urban[illegible]“[illegible]”[illegible]Urban Wallace Associates[illegible]

[illegible]“[illegible]”

[illegible]“[illegible]”[illegible]“[illegible]”[illegible]

[illegible]

[illegible]——[illegible]“[illegible]”[illegible]

[illegible]
[illegible]

[illegible]

[illegible]“[illegible]”[illegible]
[illegible]
[illegible]Amazon[illegible]Mechanical Turk[illegible]
[illegible]

[illegible]

[illegible]
Web[illegible]

[illegible]

[illegible]Web[illegible]——[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]“[illegible]”[illegible]

□□□
□□□
□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□
□□□
□□□□□□□□

□□□□□

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□“□□”□□□□□
□□□
□□□
□□□□□□□□□□□□□□□□□□□□□□□□□——□□□□□□□□□□□□□□
□□□□□□□□□□□□□□□□□□□□□□□□□Web□□□□□□□□□□□□□
□□□
□□□□□□□□□□□□□□□□□□□□□□□□□□□□

Web□□□□□□□

□□□□□□□□□□□□□□□□□□□□□□□□□□“□□”□□□□□□□□□□
□□□□□□□□□□□□□□□□□□□□□55□75□□□□□□□□□□□□□□□
□□

□□□Sans Serif□□□Arial□□□□□□□□□□Helvetica)□□□□
100%□Web□□□□□□□□□□□□□□□□□□□□□□20□□(pxel)□14□
□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□


URBAN WALLACE ASSOCIATES
Take this one-page survey and let your opinion be heard.
为了易读性，
标题以
Arial 20号
字体显示
正文以
Arial 14号
字体显示
编号也把文本
分成易读的分块
空白总是受人欢迎，
它使页面看起来亮堂，
不会让人感到压抑

[illegible]800[illegible]

[illegible]“[illegible]”[illegible]Web[illegible]IE6[illegible]

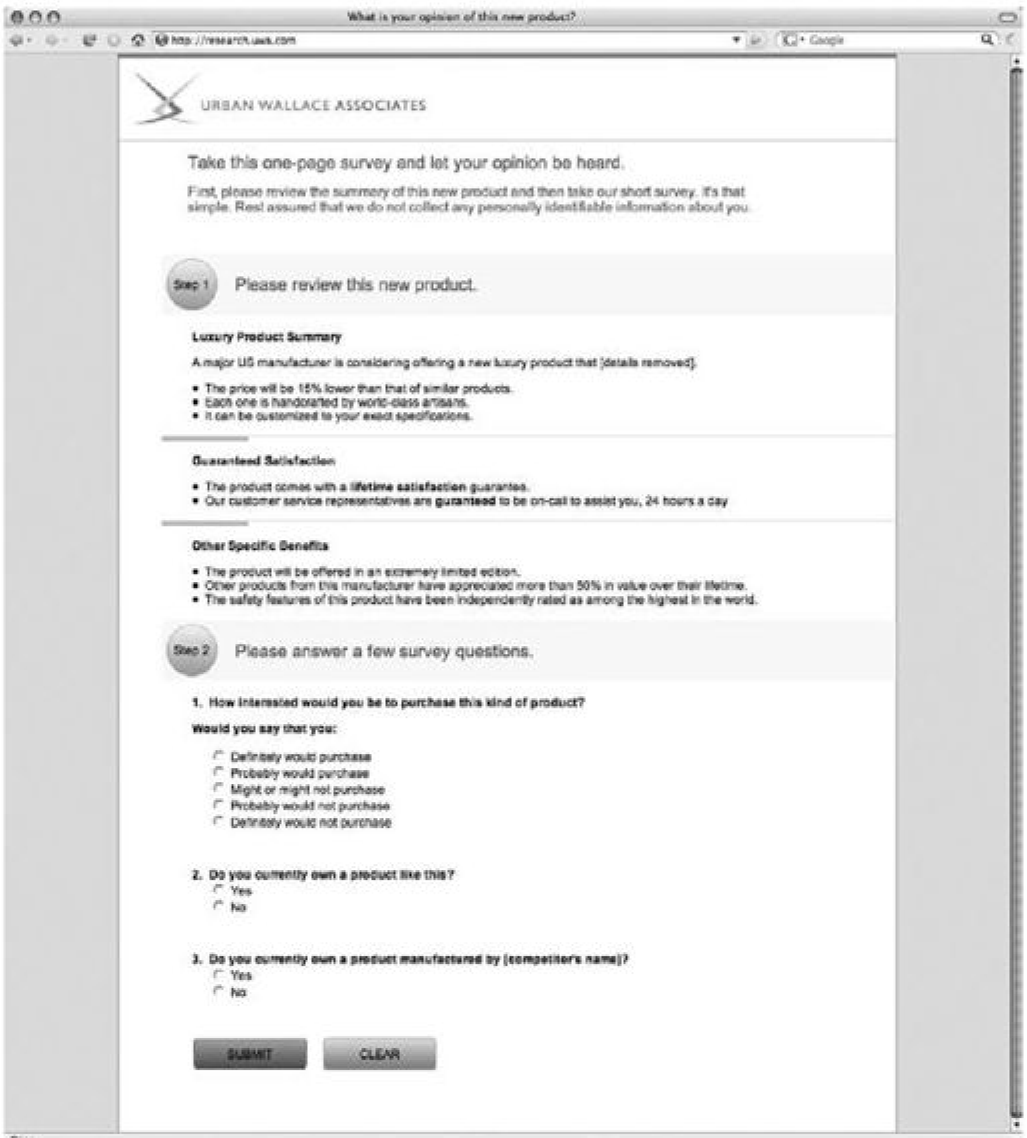
[illegible]

[illegible]“[illegible]”[illegible]“[illegible]”[illegible]JavaScript[illegible]2-3[illegible]2-4[illegible]“[illegible]”[illegible]

[illegible]JavaScript[illegible]Web[illegible]JavaScript[illegible]

```
□script language="JavaScript"□
//This finds the word"Yes"in an input value and displays the designated
hiddenElement
(o hides it if"Yes"is not found)
function switchem(switchElement,hiddenElement){
if(switchElement.value.search□"Yes"□□-1□
document.getElementById(hiddenElement).style.display=''□
else
document.getElementById(hiddenElement).style.display='none'□
}
□/script□
□script language="JavaScript"□
//This finds the word"No"in an input value and displays the designated
```

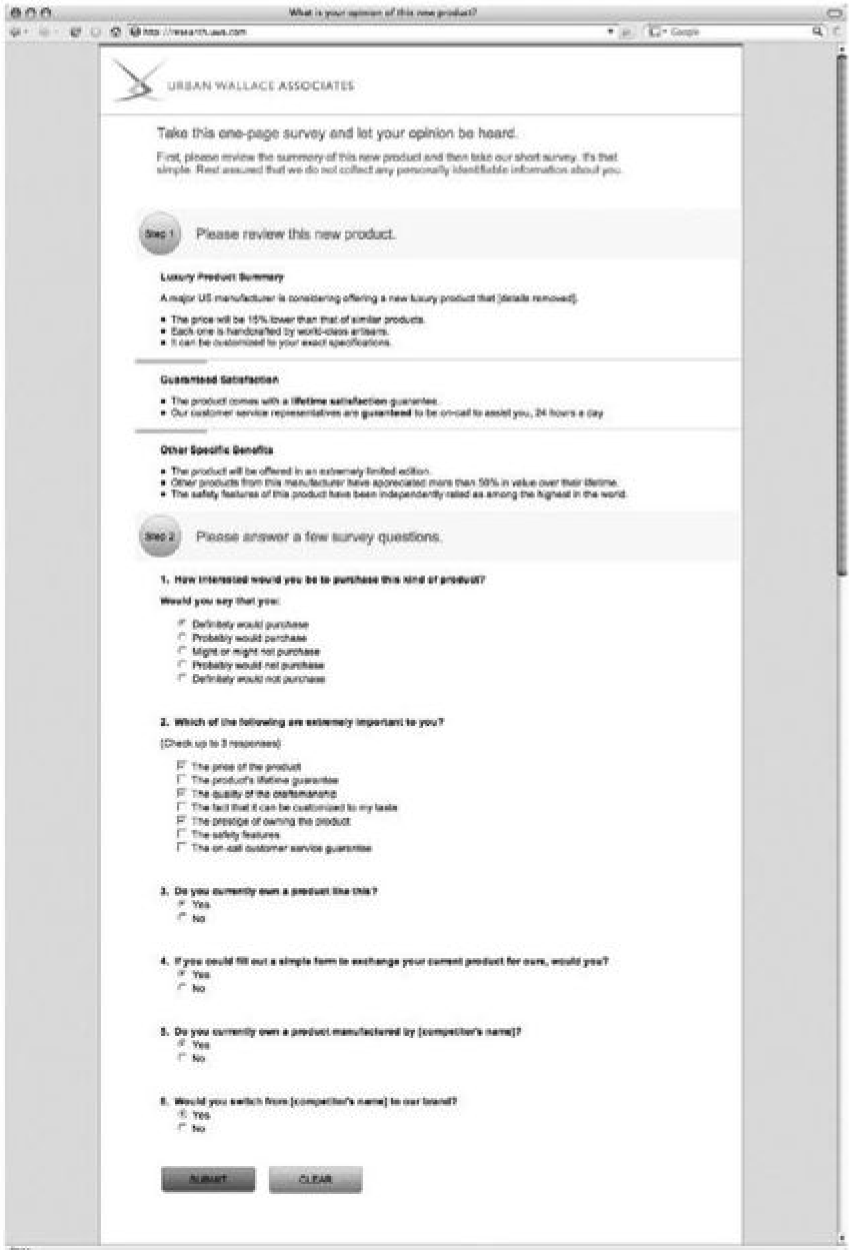
```
hiddenElement
(o hides it if"No"is not found)
function switchem2(sitchElement,hiddenElement){
if(switchElement.value.search□"No"□□-1□
document.getElementById(hiddenElement).style.display=''□
else
document.getElementById(hiddenElement).style.display='none'□
}
□/script□
```



□ 2-3□□□□□□□3□□□□□□6□

……

```
    □li id="survey1"class="surveynum"□How interested would you be to
purchase this
    kind of product?
    □p□□b□Would you say that you□□/b□□/p□
    □ul class="nobullet"□
    □li□□input
    onclick="switchem(this□'survey2'□□switchem2(tis□'survey3'□□document.
    get Element ById
□'surveytextarea'□.value=''"type="radio"name="q1"value="Yes□
    Definitely would purchase"□Definitely would purchase□/li□
    □li□□input
    onclick="switchem(this□'survey2'□□switchem2(tis□'survey3'□□document.
    getElementById
□'surveytextarea'□.value=''"type="radio"name="q1"value="Yes,Probably
    would purchase"□Probably would purchase□/li□
    □li□□input
    onclick="switchem(this□'survey2'□□switchem2(tis□'survey3'□□document.
    getElementById
□'surveytextarea'□.value=''"type="radio"name="q1"value="Yes,Might or
    might not purchase"□Might or might not purchase□/li□
    □li□□input
```



□ 2-4□□□□□□□□□□□□□□□□6□□□□□□7□

```
onclick="switchem(this□'survey2'□□switchem2(tis□'survey3'□□document.
getElementById□'q2a'□.checked=false□document.getElementById□'q2b'□.
checked=false□document.getElementById□'q2c'□.checked=false□document.
getElementById□'q2d'□.checked=false□document.getElementById□'q2e'□.
checked=false□document.getElementById□'q2f'□.checked=false□
document.getElementById
□'q2g'□.checked=false"type="radio"name="q1"value="No□
Probably would not purchase"□Probably would not purchase□/li□
□li□□input
```

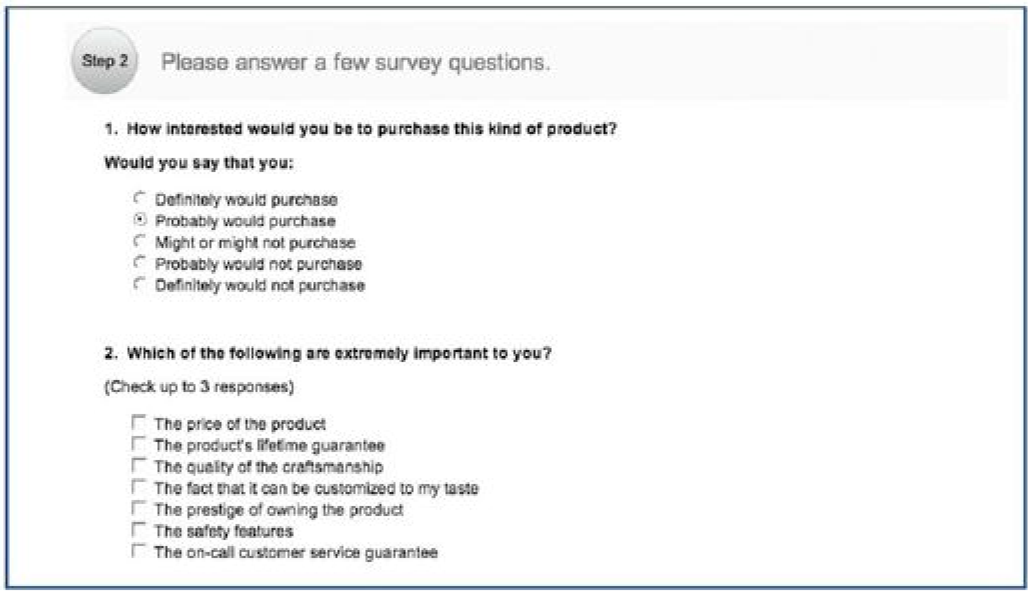
```
    onclick="switchem(this□'survey2'□□switchem2(tis□'survey3'□□document.
    getElementById□'q2a'□.checked=false□document.getElementById□'q2b'□.
    checked=false□document.getElementById□'q2c'□.checked=false□document.
    getElementById□'q2d'□.checked=false□document.getElementById□'q2e'□.
    checked=false□document.getElementById□'q2f'□.checked=false□
    document.getElementById
□'q2g'□.checked=false"type="radio"name="q1"value="No□
    Definitely would not purchase"□Definitely would not purchase□/li□
    □/ul□
    □/li□
    □li id="survey2"style="display:none"class="surveynum"□Which of the
following are
    extremely important to you?
    □p□(Ceck up to 3 responses)□/p□
    □ul class="nobullet"□
    □li□□input type="checkbox"name="q2"id="q2a"value="The price of the
product"□
    The price of the product□/li□
    □li□□input type="checkbox"name="q2"id="q2b"value="The product's
    lifetime guarantee"□The product's lifetime guarantee□/li□
    □li□□input type="checkbox"name="q2"id="q2c"value="The quality of
    the craftsmanship"□The quality of the craftsmanship□/li□
    □li□□input type="checkbox"name="q2"id="q2d"value="The fact that it can
be
    customized to my taste"□The fact that it can be customized to my taste□/li□
    □li□□input type="checkbox"name="q2"id="q2e"value="The prestige of
    owning the product"□The prestige of owning the product□/li□
    □li□□input type="checkbox"name="q2"id="q2f"value="The safety features"□
The
    safety features□/li□
    □li□□input type="checkbox"name="q2"id="q2g"value="The on-call customer
    service guarantee"□The on-call customer service guarantee□/li□
    □/ul□
    □/li□
    □li id="survey3"style="display:none"class="surveynum"□Why are you not
interested
    in this product?
    □ul class="nobullet"□
    □li□□textarea id="surveytextarea"name="q3"□□/textarea□□/li□
    □/ul□
    □/li□
```

□□□□□□5□□□□□1□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□2-5□□□□□□□□□□□□□□□□□□□□□□□□□□

[illegible]2-6[illegible]

[illegible]“[illegible]”[illegible]“[illegible]”[illegible]3[illegible]6[illegible]“[illegible]”[illegible]“[illegible]”[illegible]



Step 2 Please answer a few survey questions.

1. How interested would you be to purchase this kind of product?

Would you say that you:

- Definitely would purchase
- Probably would purchase
- Might or might not purchase
- Probably would not purchase
- Definitely would not purchase

2. Which of the following are extremely important to you?

(Check up to 3 responses)

- The price of the product
- The product's lifetime guarantee
- The quality of the craftsmanship
- The fact that it can be customized to my taste
- The prestige of owning the product
- The safety features
- The on-call customer service guarantee

[illegible] 2-5[illegible]——[illegible]1[illegible]“Yes”[illegible]8[illegible]



Step 2 Please answer a few survey questions.

1. How interested would you be to purchase this kind of product?

Would you say that you:

- Definitely would purchase
- Probably would purchase
- Might or might not purchase
- Probably would not purchase
- Definitely would not purchase

2. Why are you not interested in this product?

[illegible] 2-6[illegible]——[illegible]1[illegible]“No”[illegible]9[illegible]

[illegible]

[illegible]

[illegible]“[illegible]”[illegible]Web[illegible](Igo)[illegible]

[illegible]Urban Wallace Associates[illegible]“[illegible]”(Aout Us)[illegible]

[illegible]

[illegible]“[illegible]”[illegible]IP[illegible]IP[illegible]Web[illegible]IP[illegible](IP)[illegible]

[illegible]IP[illegible]IP[illegible]

[illegible]——[illegible]——[illegible]

[illegible]“[illegible]”[illegible]“[illegible]”[illegible]

[illegible]

[illegible]“[illegible]”[illegible]“[illegible]”[illegible]Roger Urban[illegible]“[illegible]”[illegible]“[illegible]”[illegible]“[illegible]”[illegible]

[illegible]“[illegible]”[illegible]“[illegible]”Roger Urban[illegible]

[illegible]“[illegible]”[illegible]“[illegible]”[illegible]“[illegible]”Urban[illegible]“[illegible]”[illegible]

[illegible]

[illegible]

“[illegible]‘[illegible]’”[illegible]Roger Urban[illegible]“[illegible]‘[illegible]’[illegible]‘[illegible]’”[illegible]“[illegible]”[illegible]“[illegible]”[illegible]“[illegible]”[illegible]“[illegible]”[illegible]

□□□□□□□□□□□□□□□□□□□□□□Apple□□□□i-gadget□□□□□□□□□□□□□
□□□□□□□□□

□□□

“□□□□□□□□**□□□□□□□□□□□□□□□”

“□□□□□□‘□□□□’□”

□□□“□□”□□□□□□□□□□□12.90%□□□□□□□□□□□□□□□□□□□□16.22%□□□□□□□□□□□□□□□□14.04%□□□□□□□□□□□□□□□□□□□□29.5%□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□“□□□□”□□27.68%□□□□□□□□□□□□□□□□33.16%□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□(Eail□□

□□

□□□□□□□□

□□

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□
□□□□□□□HTML□□□□□□□□□□□□□□□□□□□□□□□
□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□55□
64□□65□75□□□□□□□□□□□□□□□□□□□□□Yes/No□□□
□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□
□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□
□□□□□2-7□□



□ 2-7□□□□□□□□□□□□□□□□□□□□□

□□□□□□□□□□□□□□□□□□□□□□□□□□□□Excel□
□□□□□□□□□□□□□□□8□□□□□□□□□□□□□□□□□
□□□□□□□□□□□□□□□

[illegible]

[illegible]Web[illegible]

[illegible]——[illegible]Web[illegible]

[illegible]

[illegible]20[illegible]90[illegible]

[illegible]

[illegible]“[illegible]”[illegible]

[illegible]

[illegible]

[illegible]1%[illegible]2%[illegible]4%[illegible]

[illegible]5%[illegible]7%[illegible]21%[illegible]Web[illegible]

[illegible]2%[illegible]5%[illegible]29%[illegible]2-8[illegible]

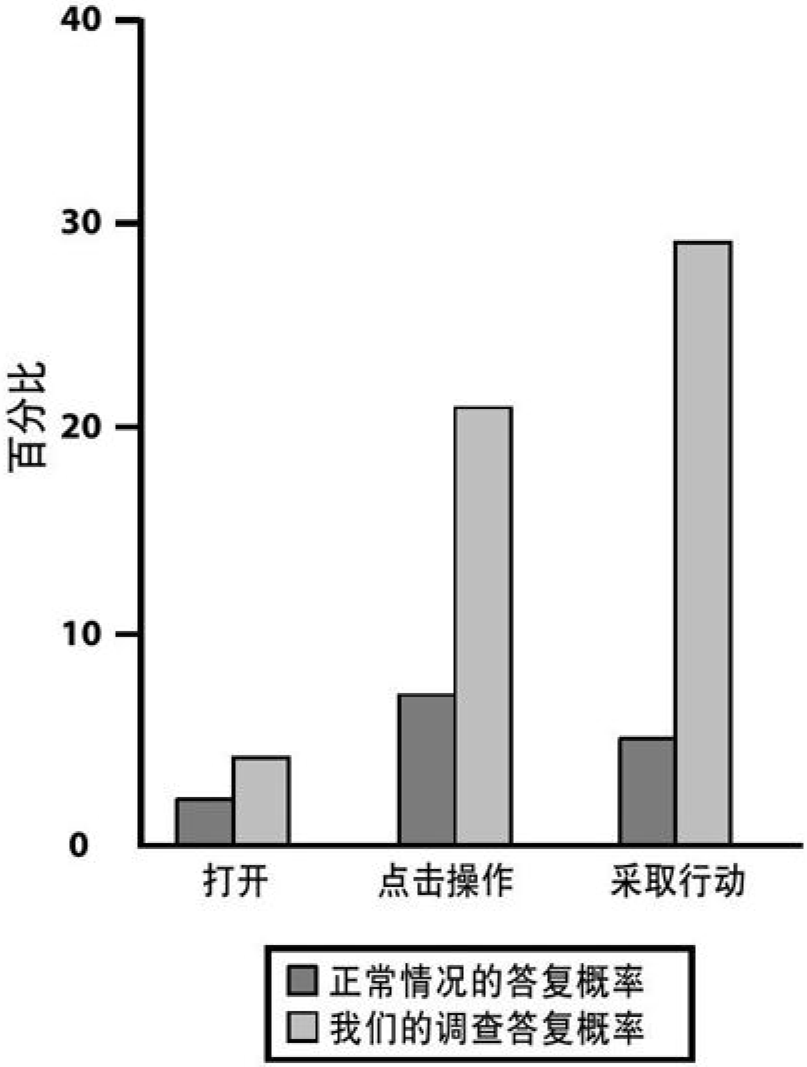


图 2-8 [illegible]

[illegible]——[illegible]

[illegible]

[illegible]10%[illegible]“Yes”[illegible]16%[illegible]

[illegible]“[illegible]”[illegible]

[illegible]3[illegible] [illegible]

J. M.Hughes

[illegible]

[illegible]“[illegible]”[illegible]“[illegible]”——[illegible]

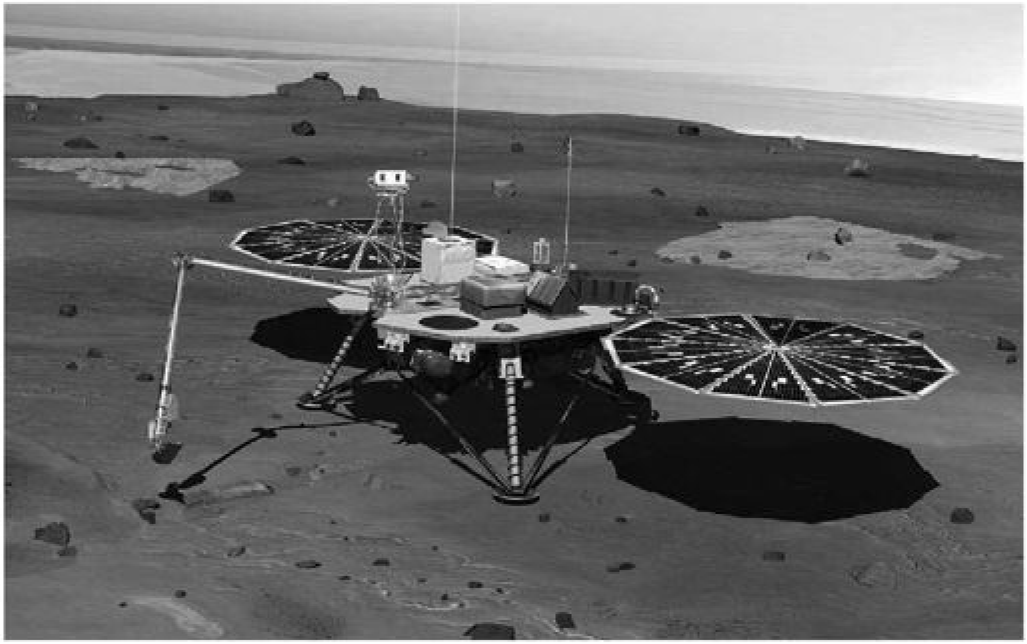
(Poenix lander)[illegible]

[illegible]

[illegible](CU)[illegible]CPU[illegible]PC[illegible]CPU[illegible]PC[illegible]

[illegible]

[illegible]2008[illegible]5[illegible]25[illegible]3-1[illegible]10[illegible]



[illegible] 3-1[illegible]
NASA/JPL[illegible]10[illegible]

[illegible]——[illegible]“[illegible]”[illegible]-90℃[illegible]

[illegible]“[illegible]”[illegible]/[illegible](JL/NASA)[illegible]“[illegible]”[illegible]

[illegible](carge-coupled device,CCD)[illegible]

25000

[illegible]

[illegible]“[illegible]”[illegible]

[illegible]CPU[illegible]RAD6000[illegible]20MHz[illegible]VME[illegible]PowerPC[illegible]CPU[illegible]“[illegible]”[illegible]WindRiver[illegible]VxWorks[illegible](ROS)[illegible]Lockheed Martin[illegible]C[illegible]

“[illegible]”[illegible]“[illegible]”

[illegible]“[illegible]”[illegible]“[illegible]”[illegible]“[illegible]”[illegible]“[illegible]”[illegible]

[illegible]“Push To Talk”[illegible]“[illegible]”[illegible]“[illegible]”[illegible]“[illegible]”(rdiated)[illegible]NASA[illegible]Deep Space Network(DSN)[illegible]“[illegible]”[illegible]

“□□”□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

□□□□□NASA□□□□DSN□□□□□□□□□□□□□□□□□□□□□□□□□

□□□□□JPL□□□□□□□JPL□Mission□□□□□□□□□□□□□□□□□□

Arizona□□□□□□□□□□□□□□□□□□□□□

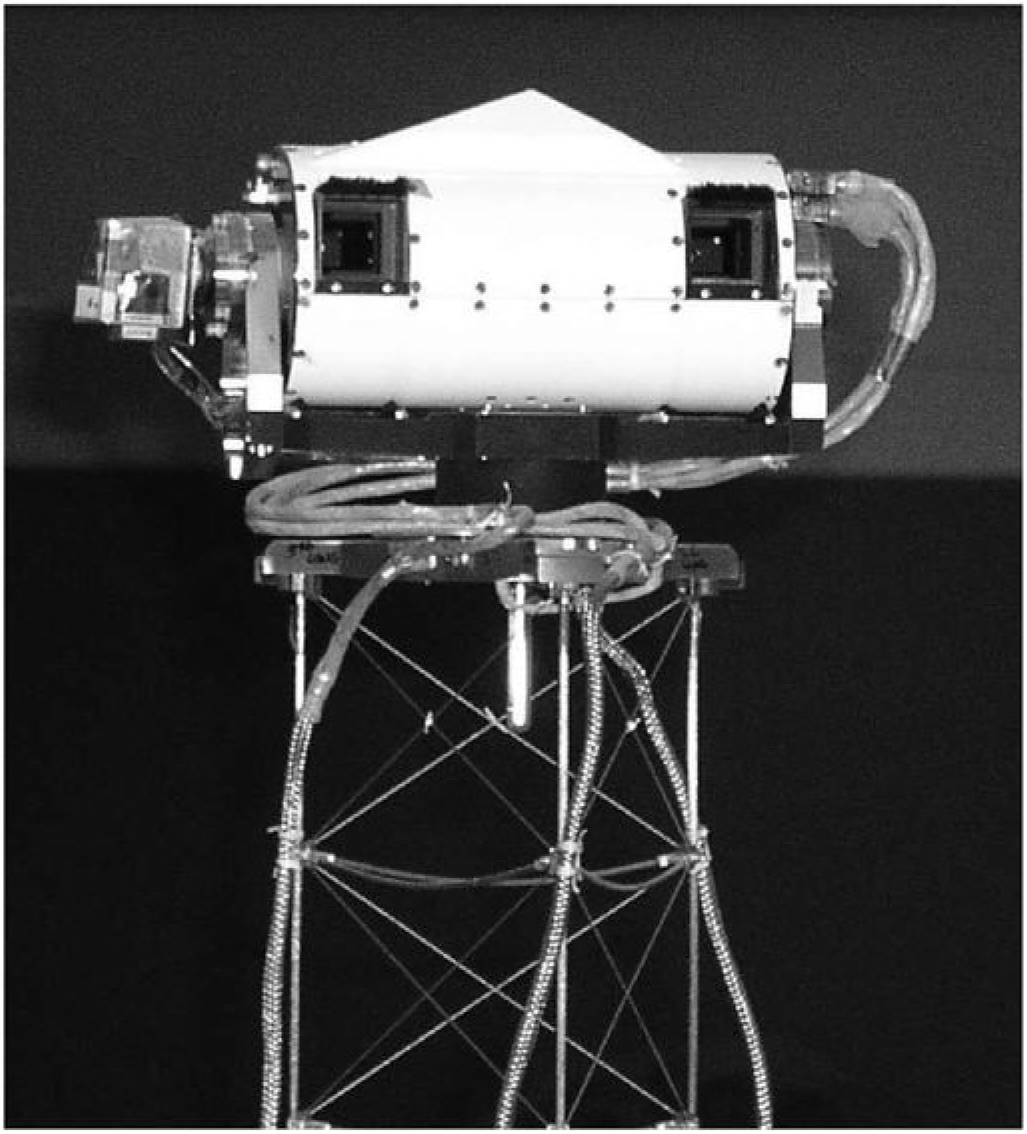
□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□(Sereo Surface Imager,SSI□□□CCD□□□□□□□□□□(Rbotic Arm Camera,RAC□□□□CCD□□□□□□□□□□□(MCA Optical Microscope,OM)□□□□□□□□□RAC□□CCD)□3-2□□□□□□□□SSI□□□□□□3-3□□□□□□□□□□□□□RAC□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

□□(sice-and-dice)□□□□□□□□□□□□□□□□□□□□□“□□”(dwnlink)□□□□□□□□□□□

□□□□□□SSI□□1024×1024□□CCD□□□□□□□2MB□□□□□□100□□12□□□□□□□SSI□□□□□□□□□□□□□□□□□“□□”□□□SSI□□□□□□□□□□□□□□□□□□□RAC□□□□□□□□□□OM□□□□□512×256□□CCD□□□□□□□□□□□□131072□□262144□□□□□□□□□□□□□□□□□□□□□□□□□RAC/OM□□□□□□□□□□□□□□□□□□□□□□□□CCD□□□□□□□□□12□□□□□□□□□□“□”□16□□□□□□□4□□□□□□□□□□□□□□□□□□□□□□□□□□□□□SSI□RAC□

OM[illegible]

[illegible]



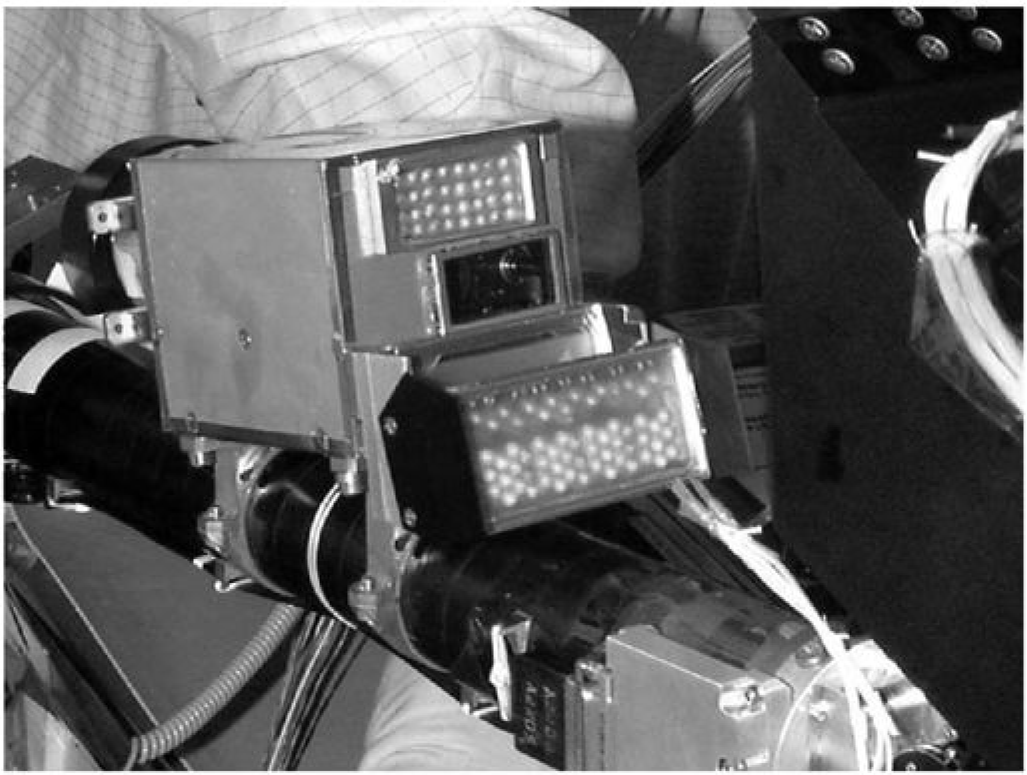
[illegible] 3-2[illegible]Arizona[illegible]/NASA/JPL[illegible]11[illegible]

[illegible]OM CCD[illegible](MCA)[illegible]

[illegible]JPL[illegible]MECA[illegible]

[illegible]

[illegible]100[illegible]
CCD[illegible]CCD[illegible]
[illegible]CCD[illegible]
[illegible](pxel-by-pixel)[illegible]CCD[illegible]
[illegible]CCD[illegible]
[illegible]CCD[illegible]



[illegible] 3-3[illegible](Rbotic Arm Camera[illegible]Arizona[illegible]/Max Planck/NASA/JPL)

[illegible]

[illegible]SSI[illegible]RAC/OM[illegible]12[illegible]“[illegible]”[illegible]

[illegible]12[illegible]12[illegible]16[illegible]“[illegible]”[illegible]12[illegible]“[illegible]”[illegible]——[illegible]16[illegible]RAD6000[illegible]20MHz[illegible]CPU[illegible]

[illegible]“[illegible]”[illegible]12[illegible]16[illegible]CPU[illegible]

[illegible]

[illegible]CPU[illegible]

[illegible]

[illegible]1[illegible]0[illegible]1[illegible]100[illegible]8[illegible]8[illegible]0[illegible]255[illegible]$11001000_2=200_{10}$[illegible]16[illegible]32[illegible]CPU[illegible]16[illegible]16[illegible]8[illegible]8[illegible]16[illegible]16[illegible]CCD[illegible]12[illegible]16[illegible]32[illegible]

[illegible]

[illegible]VxWorks[illegible]I/O[illegible]/[illegible]

[illegible]SSI[illegible]RAC/OM[illegible]SSI[illegible]Mars Rovers[illegible]CCD[illegible]RAC[illegible]OM[illegible]Max Planck[illegible]Huygens[illegible]Titan[illegible]RAC/OM[illegible]Mars'98[illegible]“[illegible]”[illegible]“[illegible]”[illegible]12[illegible]

[illegible](Lckheed Mar-tin)[illegible]“[illegible]”[illegible]SSI[illegible]RAC[illegible]OM[illegible]

[illegible]

[illegible]“[illegible]”[illegible]SSI[illegible]RAC/OM[illegible]

[illegible]“[illegible]”[illegible]

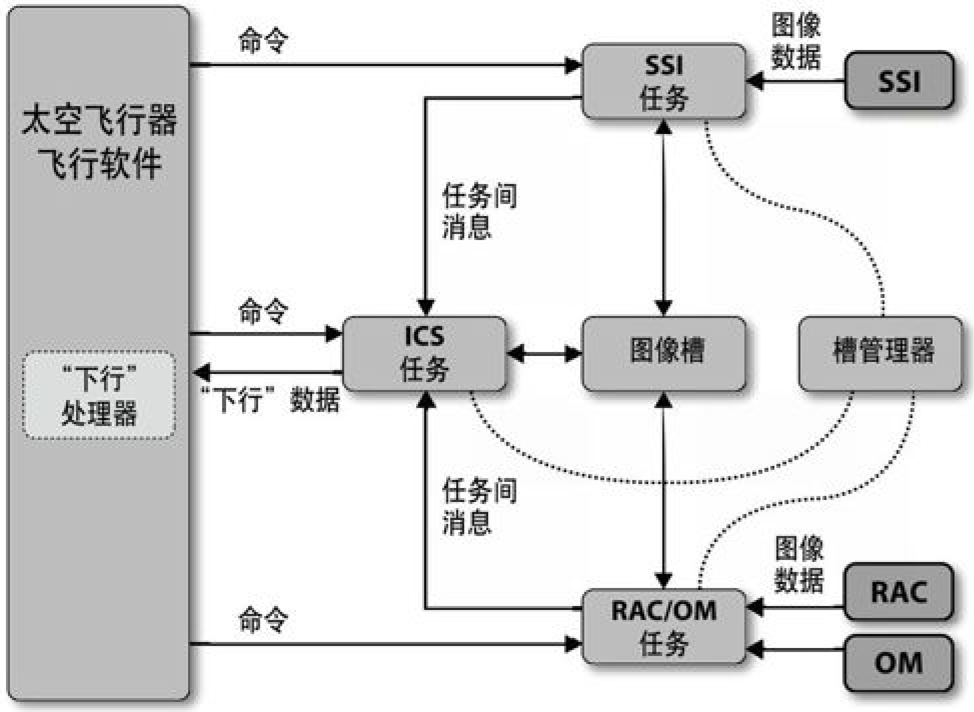
[illegible]ICS,ICS[illegible](Iage Compression Sub-System)[illegible]3-4[illegible]

[illegible]ICS[illegible]

[illegible]15000[illegible]ICS[illegible]

ICS[illegible]“[illegible]”[illegible]ICS[illegible]

ICS[illegible]ICS[illegible]



太空飞行器
飞行软件
“下行”
处理器
命令
命令
命令
“下行”数据
SSI
任务
ICS
任务
RAC/OM
任务
任务间
消息
任务间
消息
图像槽
槽管理器
图像
数据
图像
数据
SSI
RAC
OM

图 3-4[illegible]

[illegible]

[illegible]

[illegible]500M[illegible]1G[illegible]SSI[illegible]

RAC/OM[illegible]230K[illegible]10M[illegible]10255360[illegible]

[illegible]

[illegible]SSI[illegible]“[illegible]”[illegible]

[illegible]RAC/OM[illegible]

[illegible]“[illegible]”[illegible]SSI[illegible]

[illegible]

[illegible]C[illegible]malloc

[illegible]

[illegible]

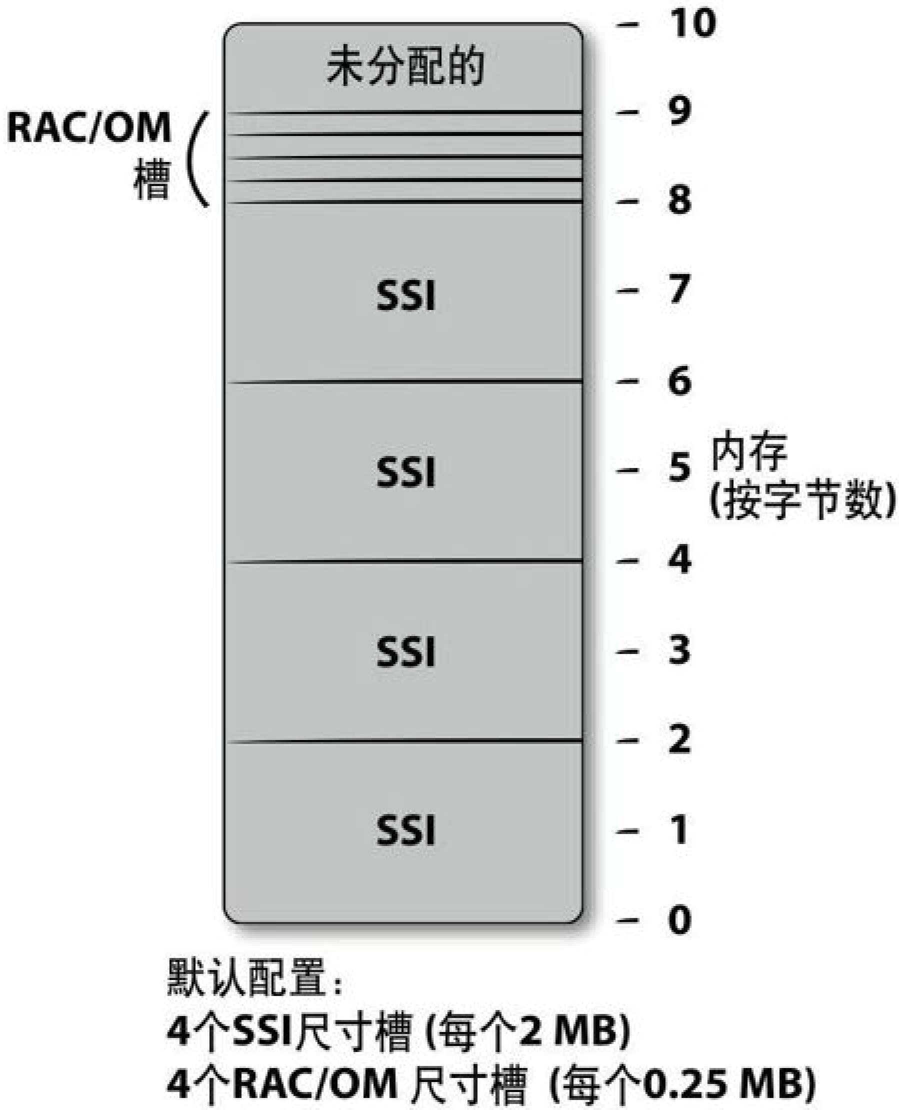
[illegible]ICS[illegible]

[illegible]

[illegible]

[illegible]“[illegible]”[illegible]SSI[illegible]

RAC/OM[illegible]3-5[illegible]ICS[illegible]


未分配的
RAC/OM
槽
SSI
SSI
SSI
SSI
10
9
8
7
6
5
4
3
2
1
0
内存
(按字节数)
默认配置:
4个SSI尺寸槽 (每个2 MB)
4个RAC/OM 尺寸槽 (每个0.25 MB)

[illegible] 3-5[illegible]“[illegible]”[illegible]

[illegible]

[illegible]

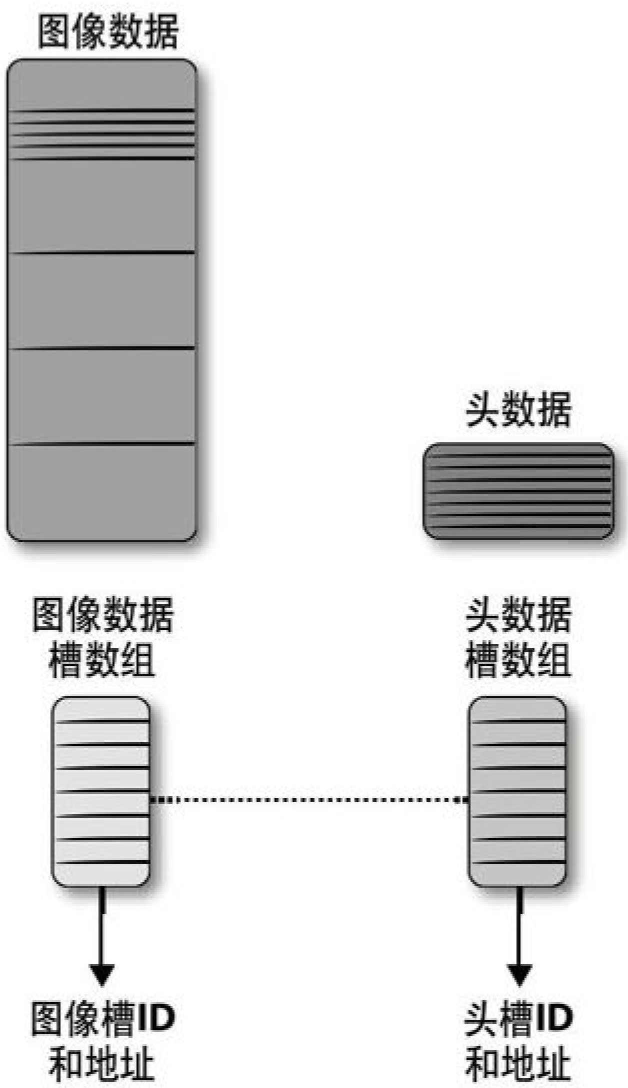
[illegible]ICS[illegible]“[illegible]”[illegible]

[illegible]“[illegible]”[illegible]

[illegible]

[illegible]3-6[illegible]

[illegible]



[illegible] 3-6[illegible]

[illegible]C[illegible]

```
typedef struct{
```

```
uint16_t slot_status□   /**□Owned or unowned   */
int16_t slot_owner□   /**□-1 if slot is unowned   */
int16_t slot_size□   /**□either RAC/OM or SSI sized   */
uint16_t*slot_address□   /**□address of data space of slot   */
}ics_img_slot_entry_t□
```

□□□□□□□□□□□□□□□□□□□□

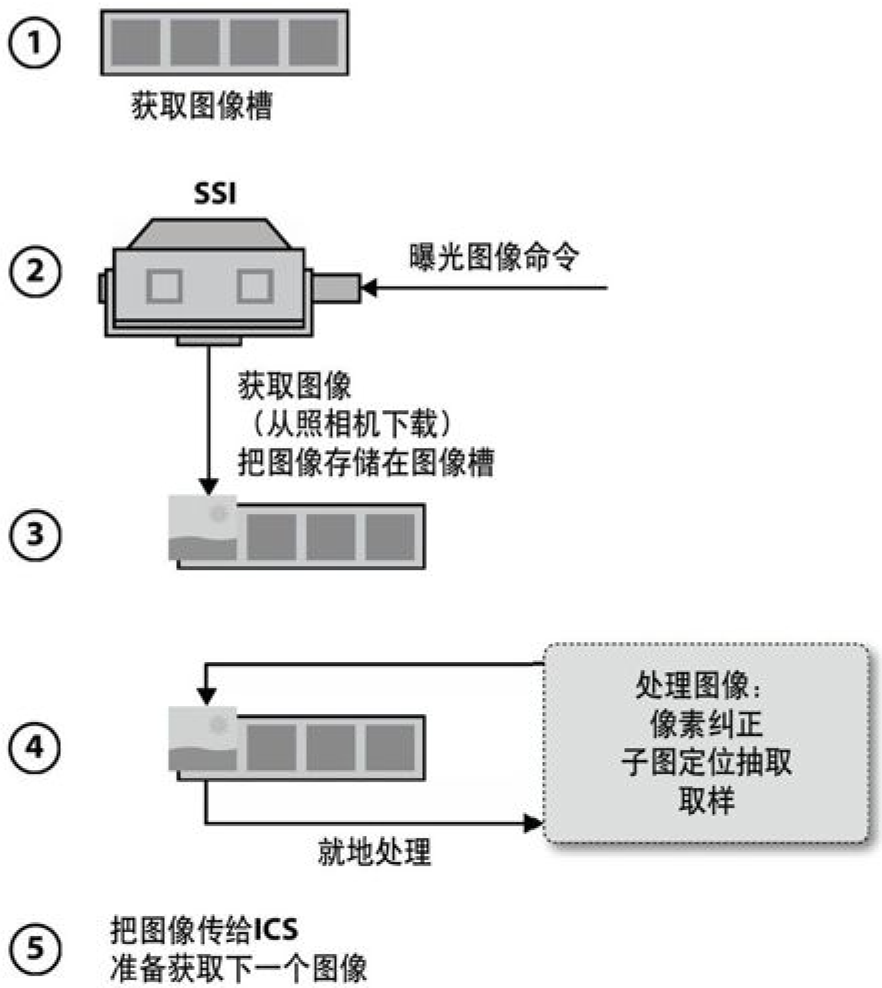
```
typedef struct{
uint16_t slot_status□   /**□Owned or unowned   */
int16_t slot_owner□   /**□-1 if unowned   */
int16_t img_id□   /**□associated image data slot number   */
uint16_t hdr_data[ICS_HDR_SLOT_SZ]□/**□array for header data   */
}ics_hdr_slot_entry_t□
```

□□□□□□□□□□□□ID□□ID□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

□□□Sol)□□SSI□□□□□□□□□□□□□□□□□RAC/MECA□□□□□□□□□□□□□□□□□□□□□□□□RAC□OM□□□□□□□□□□□□□□□□□□□□□□□39□□□□□□□□

[illegible]

[illegible]ICS[illegible]ICS[illegible]“[illegible]”[illegible]“[illegible]”[illegible]SSI[illegible]3-7[illegible]



[illegible] 3-7[illegible]

[illegible]3-6[illegible]ICS[illegible]“[illegible]”[illegible]ICS[illegible]

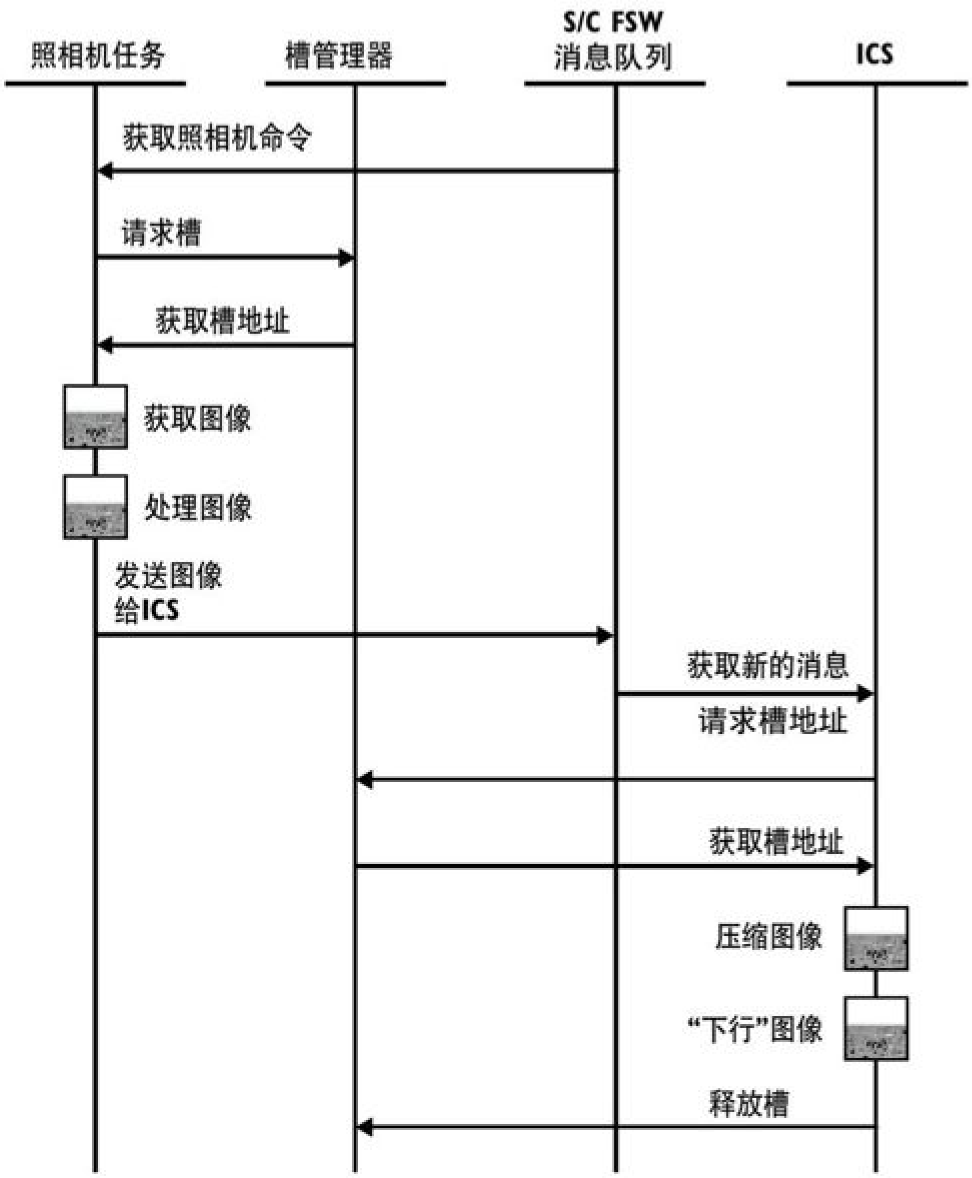
[illegible]ICS[illegible]“[illegible]”[illegible]ICS[illegible]ICS[illegible]ID[illegible]ID[illegible]ID[illegible]ID[illegible]ICS[illegible]ICS[illegible]ID[illegible]ID[illegible]

[illegible]ICS[illegible]VxWorks[illegible]ICS[illegible]3-8[illegible](MC)[illegible]“[illegible]”[illegible]

[illegible]S/C FSW[illegible]“[illegible]”[illegible]ICS[illegible]ICS[illegible]ICS[illegible]

[illegible]3-8[illegible](Filure Mode,Effects,and Criticality Analysis,FMECA)[illegible]

ICS[illegible](tming margin)[illegible]“[illegible]”[illegible]“[illegible]”[illegible]SSI[illegible]6[illegible]RAC[illegible]OM[illegible]30[illegible]



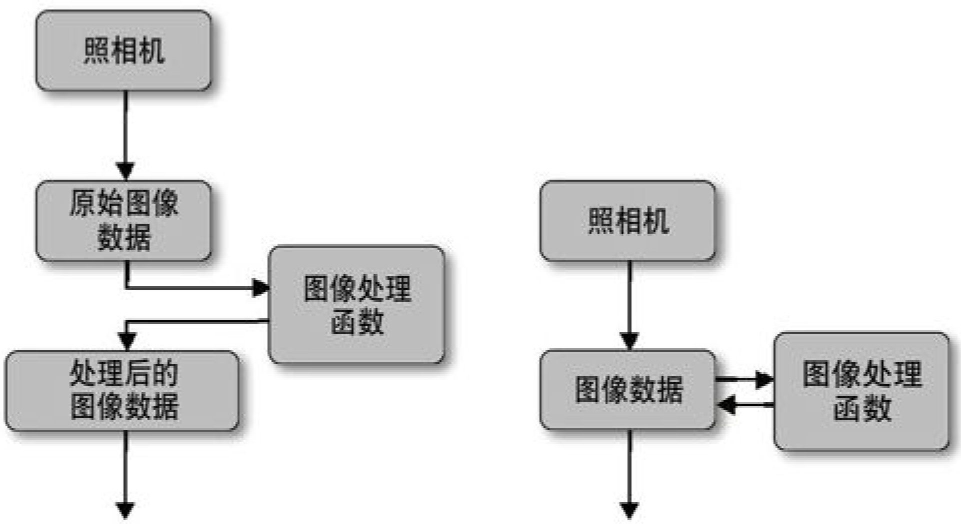
照相机任务
槽管理器
S/C FSW
消息队列
ICS
获取照相机命令
请求槽
获取槽地址
获取图像
处理图像
发送图像
给ICS
获取新的消息
请求槽地址
获取槽地址
压缩图像
“下行”图像
释放槽

[illegible] 3-8[illegible]“[illegible]”[illegible]MSC[illegible]

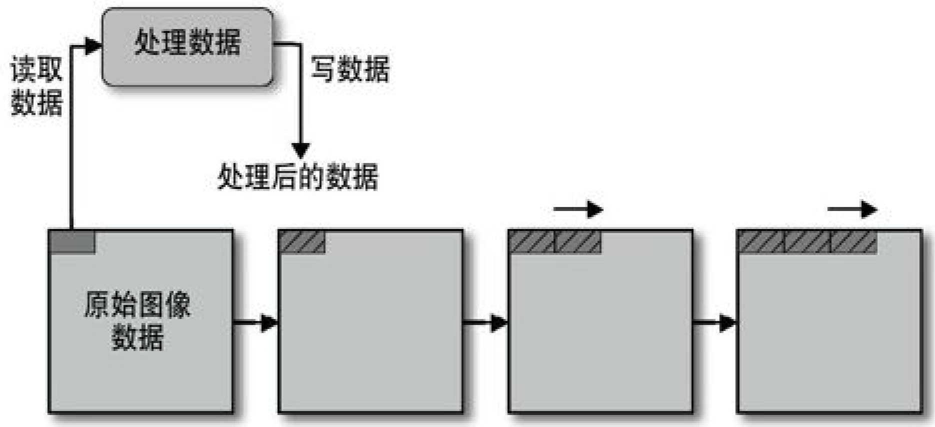
[illegible]

[illegible]3-9[illegible]

[illegible]“undo”[illegible]3-10[illegible]



[illegible] 3-9[illegible]



3-10

[illegible]

[illegible]“[illegible]”[illegible]SSI[illegible]“[illegible]”[illegible]

ICS[illegible]8[illegible]JPEG[illegible]

[illegible]ICS[illegible]JPEG[illegible]“[illegible]”[illegible]JPEG[illegible]JPEG[illegible]Mars'98[illegible]JPEG[illegible]ICS[illegible]JPEG[illegible]16[illegible]

[illegible]Rice[illegible](Rce Lossless)[illegible]Rice[illegible]Jet Propulsion[illegible]Robert Rice[illegible]Rice[illegible]2[illegible]1[illegible]JPEG[illegible]Rice[illegible]

□□□□□□□□□□□□□□□□□□□□□□□12□□□□□□□□8□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□4□□□□□□□□□□8□□□□□JPEG□Rice□□□□□□□12□□□8□□□□□□□

□□□□□□□JPEG□□“□□”□□□□□□□□□□□□□□□□□□□□□□RAD6000□□□□□□20MHz□□□□□1M□□□□□□□□□□□□□1□□□□

□□□JPEG□□□ICS□□□□□□“□□”□□□□□□□□□□JPEG□□“□□□”(iage entropy)□“□□”(qick-look)□□□□□□□□□“□□”□□□Martian□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

JPEG□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□JPEG□□□□□□□□□□□□□□□□□□□“□□”□□□□□□□□□□□□□□□□□□□□□□JPEG□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

Rice□□□□□□□□□□□□□□□□□□□□“□□”□□□“□□”□□□

“[illegible]”[illegible]

[illegible]“[illegible]”[illegible]“[illegible]”[illegible]ICS[illegible]

[illegible]JPEG[illegible]ID[illegible](ed-to-end)[illegible]“[illegible]”[illegible]

[illegible]“[illegible]”[illegible]JPL[illegible]0[illegible]“[illegible]”[illegible]

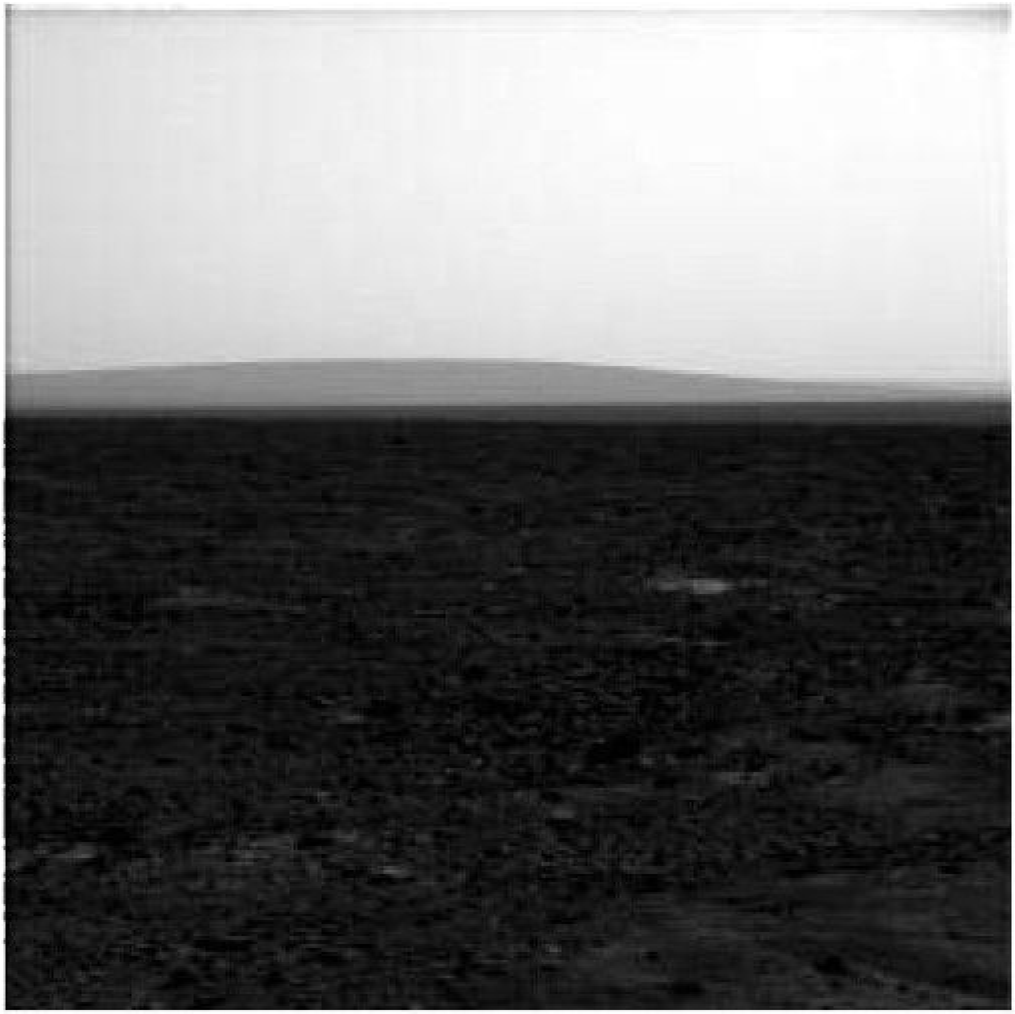
[illegible]“[illegible]”[illegible]ICS[illegible]“[illegible]”[illegible]3[illegible]10[illegible]CPU[illegible]

[illegible]

[illegible]SSI[illegible]
RAC[illegible]RAC[illegible]LED[illegible]
[illegible]SSI[illegible]RAC[illegible]
[illegible]“[illegible]”[illegible]
[illegible]
[illegible]CPU[illegible]“[illegible]”[illegible]
[illegible]

[illegible]“[illegible]”[illegible]“[illegible]”[illegible]
[illegible]“[illegible]”[illegible]
[illegible]“ping”[illegible]
[illegible]ping[illegible]
[illegible]

[illegible]SSI[illegible]RAC/OM[illegible]
[illegible]3-11[illegible]Sol
1[illegible]SSI[illegible](S001EDN896308958_10D28R1M1[illegible]



□ 3-11□□Sol 1□SSI□□□□□□□□□□□□NASA/JPL/Arizona□□□

□□□□□□□□□□□□□□□

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

Arizona□□□Phoenix□□□·http□//phoenix.lpl.arizona.edu

Jet Propulsion Laboratory(JPL□□□□□Phoenix□□□

·http□//www.jpl.nasa.gov/news/phoenix/main.php

NASA□Phoenix□□□

·http□//www.nasa.gov/mission_pages/phoenix/main/index.ht

ml

□JPL,MIPL□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

JPL□Mission□□□□□□□□·http□//www-mipl.jpl.nasa.gov/

□□□□□RAD6000 CPU□□□□□□□□□□□□□□□□□□□□□

Wikipedia:http□//www.wikipedia.org□

□4□ PNUTShell□□□□□□

Brian F. Cooper□Raghu Ramakrishnan□Utkarsh Srivastava

[illegible]

[illegible]5[illegible]Web[illegible]

[illegible]

PNUTS[illegible]Web[illegible](Coper[illegible]2008[illegible](tnant applications)[illegible]PNUTS[illegible]

[illegible](sale-out)

[illegible]

[illegible](go-replication)

[illegible]PNUTS[illegible][1](clos[illegible]

PNUTS[illegible]API[illegible]

[illegible](Rmakrishnan[illegible]Gehrke 2002[illegible]

[illegible]1[illegible]

[illegible](mterialized views)[illegible]

PNUTS

[1]

□□□□

□□□Web□□

□□□□

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□Alice□□Alice□□□Alice□□□□□□□□□□“□□□□□□”(Bsy on the phone)□□□□□□“□□□□□□□”(Of the phone,anybody wanna chat?□□□Alice□□4-1□□□□□□□Web□□□

表4-1：用户信息表

Username	FullName	Location	Status	IM	BlogID	Photo	…
Alice	Alice Smith	Sunnyvale, CA	Off the phone, anybody wanna chat?	Alice345			…
Bob	Bob Jones	Singapore	Eating dinner		3411	*me.jpg*	…
Charles	Charles Adams	New York, New York	Sleeping		5539		…
…							

□□□□□□□Alice□□□ACID[1]□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□ACID□□□□□□□□□G□□□BigTable(Chang□2006□□□Alice□□□□□□□□□□□“OK”□□Alice□□□□□□□□□□□□□□□□□□□□□□□□□□□Alice□□□□

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□Web□□□□□□□“□□□□□□”□□□(bst-effort approach)□□□□□□□□□□□□□□□□□□□□□□□□□□□(popagate)□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□——“□□□□□”□□Alice□□□□□□□□□□□“Busy”□□□□□□□□□□(wst coast)□□□□□□□□□□□□□□□□4-2□□□

表4-2：在美国西海岸的数据副本上执行的更新操作

美国西海岸	
用户名	状态
Alice	Busy

美国东海岸	
用户名	状态
Alice	--

□□□Alice□□□□□□□□□“Off the phone”□□□□□□□□□□□□□□□□□□□□□□□(est coast)□□□□□□□□□□□□4-3□□□

表4-3：在东海岸数据副本上执行的第二次更新操作

美国西海岸	
用户名	状态
Alice	Busy

美国东海岸	
用户名	状态
Alice	Off the phone

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□“Off the phone”□□□□□□□□“Busy”□□□□□□□□□□□□□□□□□□□4-4□□□

表4-4：传播中的两个更新交叉

美国西海岸	
用户名	状态
Alice	Busy



美国东海岸	
用户名	状态
Alice	Off the phone

□□□□□□“Busy”□□□□□“Off the phone”□□□□□□□□□□“Off the phone”□□□□“Busy”□□□□□□□4-5□□□□□□□□□

表4-5：不一致的数据副本

美国西海岸	
用户名	状态
Alice	Off the phone

美国东海岸	
用户名	状态
Alice	Busy

□□□□□□□□□□□Alice□□□□□Alice□□□□□□□□□□□□□□□□□□□□□Alice□□□□□□□□

□□□□□□□□□□Web□□(wb-scale)□□□□□□□□□□□□□(eentual consistency)□□□Amazon□Web□□S3□□□□□□□□□□□□□□(gssip)□□□(ati-entropy)□□Alice□□□□□□□“Busy”□□“Off the phone”□□□□□□□□□□“Busy”□□□□□□□Alice□□□□□□□□□□□□□□□□□□□□□□□Alice□□□□□□□□□□□□□□□□□□□□□

□□□□□

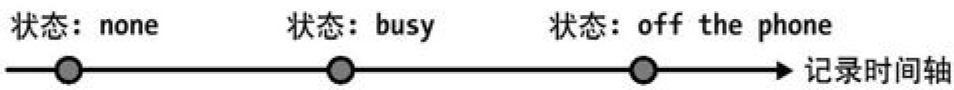
□□□□□□□□ACID□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□(tmeline

consistency)□□

□□□□□□□□□□□□4-1□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

□□□□□□□□□□□□□□□□□□□□□



□ 4-1□Alice□□□□□□□□□□

□□□□□□□□□□□□□□□□□□□□□□(mster copy□□□□□□□□□□□□□□□□

□□

□□

□□

□□□□□□□□□□□□□□□□□□(uage patterns)□□□□□□□□□□□□

(mstership)□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□Alice□□□□□□

□□□□□□□□□□□□□□□□□□□□□□□□□□□□Bob□□□□□□□□□□□□□□□□□□□

□□□□□□□□□□□□□□□□□□□□(mtadata field)□□□□□□□□□□□4-6□□□

表4-6：包含主备份性和版本元数据的信息表

Username	_MASTER	_VERSION	FullName	…
Alice	West	32	Alice Smith	…
Bob	Asia	18	Bob Jones	…
Charles	East	15	Charles Adams	…
…				

[illegible]——[illegible]Alice[illegible]
[illegible]
[illegible]Alice[illegible]
[illegible]
[illegible]85%[illegible]Alice[illegible]
[illegible]
[illegible](acess pattern)[illegible]
[illegible]

[illegible]
[illegible]
[illegible](u-to-date read)[illegible]
[illegible]
[illegible]Web[illegible]
[illegible]Alice[illegible]Bob[illegible]
[illegible]

[illegible]“[illegible]”(citical read)[illegible]
[illegible]Alice[illegible]Bob[illegible]Alice[illegible]
[illegible]Alice[illegible]Bob[illegible]
[illegible]Alice[illegible]Bob[illegible]
[illegible]
[illegible]Bob[illegible](sssion state)[illegible]cookie[illegible]

Bob Alice “ ”

Alice “ ”

“ ”(tst-and-set) “ ” ACID (otimistic concurrency control)

(rcord-level)

(oerride)

[illegible]PNUTS[illegible](lvel of indirection)[illegible](uinterrupted system availability)[illegible]

[illegible]

PNUTS[illegible]YDHT[illegible](Yhoo!Distributed Hash Table)[illegible]YDOT[illegible](Yhoo!Distributed Ordered Table)[illegible](ky space)[illegible]“[illegible]”(htpot)[illegible]“[illegible]”[illegible]

[illegible]“[illegible]”[illegible](ht partition)[illegible][2][illegible]

(ky space)

(Slberstein 2008

21 30

(ierator)

(prtition tolerance) (tade-off)

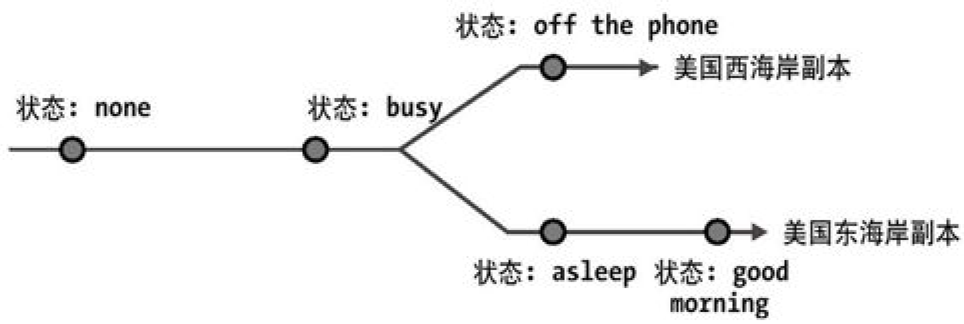
(pr-table basis)

4-2

[illegible]

[illegible]

[illegible]“[illegible]”[illegible](cmmutative)[illegible]



[illegible] 4-2[illegible](frks)

PNUTS[illegible]Web[illegible]Alice[illegible]Bob[illegible]Alice[illegible]Bob[illegible]Bob[illegible]Alice[illegible]Alice[illegible]Bob[illegible]ACID[illegible](bndled

writes)[illegible]

[illegible][illegible](smantics)[illegible]

[illegible](crrency)[illegible]ACID[illegible]

[1][illegible]

[2][illegible]

[illegible]

[illegible]Web[illegible]

[illegible]

[illegible]

[illegible]Alice[illegible]

[illegible]Alice[illegible]Alice[illegible]ID[illegible]4-7[illegible]

表4-7：好友关系表

User1	User2	…
Alice	Bob	…
Alice	Charles	…
Alice	Dave	…
…		

[illegible]“Alice”[illegible]

[illegible]

[illegible]Alice[illegible]Bob[illegible]Charles[illegible]Dave[illegible]

[illegible]Alice[illegible](PotoID,CommentID)[illegible]YDOT[illegible]4-8[illegible]

表4-8：照片评论表

PhotoID	CommentID	Comment	Commenter
Photo123	18	Cool	Mary
Photo123	22	Pretty	Alice
Photo123	29	Interesting	Charles
…			

[illegible]Alice[illegible]Alice[illegible]username[illegible]username[illegible]

[illegible](goup-by-aggregate)[illegible]Alice[illegible]Alice[illegible]

[illegible]

[illegible]

[illegible]Alice[illegible]

[illegible](Aarawal[illegible]2009[illegible]Alice[illegible][1][illegible]

[illegible]Alice[illegible]username)[illegible]Alice[illegible]Alice[illegible]username[illegible]Alice[illegible]/[illegible]“Alice”[illegible]4-9[illegible]

表4-9：信息和评论记录的连接和聚合

Alice	West	32	Alice Smith	…	← 信息记录
Alice	Photo123	22	Pretty	…	← 评论记录
Alice	Photo203	43	Nice	…	↓
Alice	Photo418	33	OK	…	
…					

[illegible]

□□□

□□(shema)□□□□□□□□□□□□□□PNUTS□□Web□□□□□□□□□Web□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□color□weight□RAM□flavor)□□

□□□(pe-group)□□□□□□□□□(pe-ag gregate)□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□“□□”□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□Sunnyvale□□□□□□□□□□□□□□□□username□□“location”□□□□□□□□□□□□□□□□□YDOT□□□□□“Sunnyvale,California”□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□4-10□□

表4-10：地理位置索引

Location	Username
Sunnyvale, CA	Alice
Sunnyvale, CA	Mary
Sunnyvale, CA	Steve
Sunnyvale, CA	Zach
…	

[illegible]Web[illegible]

[illegible]PNUTS[illegible]

PNUTS[illegible]Hadoop[illegible]Hadoop[illegible]

MapReduce[illegible](dclarative

queries)[illegible]PNUTS[illegible]

[1][illegible]PNUTS[illegible]

[illegible]

[illegible]PNUTS[illegible]G[illegible]Amazon[illegible]

G[illegible]BigTable

BigTable(Chang[illegible]2006[illegible]G[illegible]Web[illegible]“[illegible]”(bg table)[illegible](tblet)[illegible]BigTable[illegible]

[illegible]BigTable[illegible]G[illegible]GFS[illegible](Gemawat[illegible]2003[illegible]GFS[illegible]Alice[illegible]

BigTable[illegible]GFS[illegible]PNUTS[illegible]GFS[illegible]MapReduce)[illegible]BigTable[illegible]GFS[illegible]SSTables[illegible]GFS[illegible]BigTable[illegible]

[illegible]“[illegible]”[illegible]B[illegible]

PNUTS[illegible]BigTable[illegible]PNUTS[illegible]MegaStore[illegible]BigTable[illegible](Frman[illegible]2008[illegible]API[illegible]BigTable[illegible]

Amazon[illegible]Dynamo

Dynamo(DeCandia[illegible]2007[illegible]Amazon[illegible](Dnamo[illegible]Dynamo[illegible]Gossip[illegible](lzily propagate)[illegible]

[illegible]Dynamo[illegible]PNUTS[illegible](lzily propagate updates)[illegible]PNUTS[illegible]Gossip[illegible]Alice[illegible]Gossip[illegible]Texas[illegible](dterministic approach)[illegible]Gossip[illegible]

PNUTS□Dynamo□□□□□□□□□□□□□□Gossip□□□□□□□□□□□□□

□□

□“□□”□□□□□□□Alice□□□□□“Sleeping”□□“Busy”□□□□□□□□

□“Home”□□“Work”□□□□□□□□□Alice□□□□□□□□□□□□□□

(Seeping,Home)□(Bsy,Home)□(Bsy,Work)□□□□□□□□□□□□□□

□□□□□□□□□□□□□□□□□□□□□□□□□□“Work”□□□□□□□□□□□□□

□(Seeping,Work)□□□□□Alice□□□□□□□□□Alice□□□□□□□□□□□□□

□□□

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

PNUTS□Dynamo□□□□□□□□Dynamo□□□□□□□□□□□□□

PNUTS□□□□□□Dynamo□□□PNUTS□□□□□□□□□□□□□□□□□□□□

□□□□□□□□□□□□□□□□□□□□□□□□□“□□”□□□□DynamO'Amazon

□□□□□□□□□□S3□□□□□□□SimpleDB□□□□□□□□□□□□□□□□□□

□□□□SimpleDB□□□□□□API□□□□□□□□□□□□□□□□□□□□□□□□□□

□□□□□□□□□□□□□□□□□□□□□□

□□□Azure SDS

□Azure□□□□□□□□□□□□□□□□SQL Server□□□□□SQL□□□□

(SL Data Service□SDS□□□http□//hadoop.apache.org)□SDS□□

□□□□□□□□□□□□□□□□SDS□□□□□□□□□□□□□□□□□□□SQL

Server□□□□□□□□□□□□□□□□□□SDS□□□□□□□□□□□□□□□□□□□□

□□□

[illegible]“[illegible]”[illegible]

[illegible]PNUTS[illegible]

[illegible]SDS(PNUTS[illegible]PNUTS[illegible]

[illegible]

PNUTS[illegible]SDS[illegible]PNUTS[illegible]SDS[illegible]

[illegible]

Alice[illegible]Alice[illegible]

[illegible]

[illegible]Facebook[illegible]

Cassandra[illegible](Lkshman[illegible]2008[illegible]P2P(Peer-to-Peer[illegible]

[illegible]BigTable[illegible]Dynamo[illegible]

Sharded[illegible]Flickr[Pattishall][illegible]Facebook[Sobel 2008][illegible]

MySQL[illegible](sarding)[illegible]

Sharding[illegible]SimpleDB[illegible]

Sharded[illegible]

PNUTS[illegible]

[illegible]

PNUTS[illegible]

[illegible]PNUTS[illegible]Hadoop[illegible]MapReduce[illegible](Dan[illegible]Ghemawat2007[illegible]

[illegible](htp[illegible]//hadoop.apache.org)[illegible]

Hadoop[illegible]HDFS[illegible]MapReduce[illegible]

[illegible]PNUTS[illegible]MObStor[illegible]

[illegible]MObStor[illegible]

[illegible](Coper[illegible]2009[illegible]

[illegible]

[illegible]PNUTS[illegible]ACID[illegible](crner cases)[illegible]PNUTS[illegible]

[illegible]PNUTS[illegible](mltitenanted)[illegible]

□□

PNUTS□□□□□□□□□□□□□□□□□□□□P.P.S.Narayan□Chuck Neerdaels□□□□□□□□□□Adam Silberstein□Rodrigo Fonseca□Brad McMillen□Pat Quaid□□□PNUTS□□□□□□□□□□□□□□□□□□□□□□□□□□□□□Phil Bohannon□Ramana Yerneni□Daniel Weaver□Michael Bigby□Nicholas Puz□Hans-Arno Jacobsen□Bryan Call□Andrew Feng□

□□□□

[1]Azure Services Platform. http□//www.microsoft.com/azure/.

[2]Hadoop. http□//hadoop.apache.org.

[3]Agrawal,P.□A. Silberstein,B.F.Cooper,U.Srivastava,and R.Ramakrishnan."Asynchronous View Maintenance for VLSD Databases."In SIGMOD□2009.

[4]Chang,F. et al."Bigtable:A distributed storage system for structured data."In OSDI□2006.

[5]Cooper,B. F.□ E.Baldeschwieler,R.Fonseca,J.J.Kistler,P.P.S.Narayan,Chuck Neerdaels,Toby Negrin,Raghu Ramakrishnan,Adam Silberstein,Utkarsh Srivastava,and Raymie Stata."Building a cloud for Yahoo□"IEEE Data Engineering Bulletin□32□1□□36-43□2009.

[6]Cooper,B. F.□ R.Ramakrishnan,U.Srivastava,A.Silberstein,P.Bohannon,H.-A.Jacobsen,N.Puz,D.Weaver,and R.Yerneni."PNUTS:Yahoo□'s hosted data serving platform."In VLDB□2008.

[7]Dean,J. and S.Ghemawat."MapReduce:Simplified data processing on large clusters."In OSDI，2004.

[8]DeCandia,G. et al."Dynamo:Amazon's highly available key-value store."In SOSP，2007.

[9]Furman,J. J.，J.S.Karlsson,J.-M.Leon,A.Lloyd,S.Newman and P.Zeyliger."Megastore:A Scalable Data System for User Facing Applications."In SIGMOD，2008.

[10]Ghemawat,S.，H. Gobioff,and S.-T.Leung."The Google File System."In SOSP，2003.[11]Lakshman,A.，P.Malik,and K.Ranganathan."Cassandra:A Structured Storage System on a P2P Network."In SIGMOD，2008.

[12]Pattishall,D. V."Federation at Flickr:Doing Billions of Queries Per Day."http：//www.scribd.com/doc/2592098/DVPmysqlucFeder ation-at-Flickr-Doing-Billions-of-Queries-Per-Day.

[13]Ramakrishnan,R. and J.Gehrke.Database Management Systems.McGraw-Hill,New York,NY，2002.

[14]Silberstein,A.，B. F.Cooper,U.Srivastava,E.Vee,R.Yerneni,and R.Ramakrishnan."Efficient bulk insertion into a distributed ordered table."In SIGMOD，2008.

[15]Sobel,J."Scaling out."Facebook Engineering Blog,August 2008.

[illegible]5[illegible] [illegible]

Jeff Hammerbacher

[illegible]

[illegible]17[illegible]Fort Wayne[illegible]Scott[illegible]

[illegible]10[illegible]

[illegible]17[illegible]

[illegible]

[illegible]Alex Wright[illegible]Glut[illegible](Jseph Henry[illegible]Daniel Dennett[illegible]“[illegible]”[illegible]a scholar is just a library's way of making another library)[illegible]

[illegible]2006[illegible]Facebook[illegible]Facebook[illegible](ptabyte,PB)[1][illegible]Facebook[illegible]

[1]1PB(petabyte)=1024TB，1TB(terabyte)=1024GB

Facebook[illegible]“[illegible]”

[illegible]2005[illegible]9[illegible]Facebook[illegible]Facebook[illegible]

[illegible]Facebook[illegible]Facebook[illegible]Facebook[illegible]Facebook[illegible]

[illegible]2006[illegible]2[illegible]Facebook[illegible]GPA[illegible]“[illegible]”[illegible]Facebook[illegible]

[illegible]Facebook[illegible]

[illegible]Python[illegible]Facebook[illegible]MySQL[illegible]C++[illegible]10GB[illegible]“ETL”[illegible]

Python[illegible]C++[illegible]Facebook[illegible]MySQL[illegible]MySQL[illegible]“[illegible]”(Dta Warehouse)[illegible]

□□□□□□□□PHP□□□□□□□□□MySQL□□□□□□□□□□□□□□□□□□□□□□□(smmary)□□□Email□□

□□□□□□□□□□□□□□□ETL□□□□□□□□□□□□□(dshboard)□□□□□□□□□□□□□“□□□□”(Bsiness Intelligence)□□□

□□□□□□

□1958□IBM□□□□□□□□□□Hans Peter Luhn□□□□□□□□□□□□□
□□□“□□□”(ation points)□“□□□□”(iterest profiles)□□□□□“□□□□
□”(slective dissemination)□□□□□□□□□□□□□□□□□□□□□□□□□□□
□“□□□□□□□□”□□□“□□□□”□□□□□□□□□□□□□□□□

□□□□□□□□□□□□□□□□“□□□□”□□□□□□□□□□□□□□□□Luhn□□□□
□□
□□□□□□□□□□□□□□□□□□□□□□□□□□□“□□□”□□□□□□□□□□□□□□□
□□□□□□□□□□□

□Luhn□□□□□□□□□□□□□□□□□□□□□□□□“□□□□”□□□□□□□□□□
□□□□□□□□□□□□□□□□□□□□□ETL□□□□□□□□□□□ETL□□□□□□□□
□□
□□□□□□□Luhn□□□□□□□

□1970□□E.F.Codd□□□□□□□□□□20□□70□□□□□IBM□□□□□□□
□□□□(RBMS)□□□□□RDBMS□□□□□□□□□□□□□□□□20□□80□□□
□□□□□□□□□□□□

□1983□□Teradata□□□□□□□□□□□□□□□□□□Wells Fargo□□□□
□□□□□□□□□□□1986□□Ralph Kimball□□□Red Brick Systems□
□□□□□□□□□□□□□□□□□□□□□□□□Teradata□Red Brick□□□□□□□
□□□□□1991□□□□□□□□□□□□□□□□□□

Bill Inmon[illegible]Building the Data Warehouse[illegible](Wley[illegible]Inmon[illegible]

[illegible]1995[illegible]Inmon[illegible]TDWI(The Data Warehouse Institute)[illegible]TDWI[illegible]WHIPS[illegible]

1996[illegible]Ralph Kimball[illegible]The Data Warehouse Toolkit[illegible](Wley[illegible]Inmon[illegible]Kimball[illegible](nrvana)[illegible]Kimball[illegible]“[illegible]”(mrts)[illegible]“[illegible]”(bs)[illegible]Kimball[illegible]

[illegible]“[illegible]”[illegible]Jim Gray[illegible]1997[illegible]CUBE[illegible]

[illegible]CUBE[illegible](dcument)[illegible](ation point)[illegible]Luhn[illegible]

[illegible]Facebook[illegible]

[illegible]

[illegible]Facebook[illegible]MySQL[illegible]

[illegible]100[illegible]1TB)[illegible]mysqld[illegible]

[illegible]“[illegible]”(rdo log)[illegible]“[illegible]”(wite-ahead log)[illegible]MySQL[illegible]

[illegible]Oracle[illegible]Oracle[illegible]Sun[illegible]

[illegible]MySQL[illegible]Oracle[illegible]SQL[illegible]Oracle[illegible]Python[illegible]bug[illegible]

[illegible]Oracle[illegible]Oracle“[illegible]”[illegible]Oracle[illegible]Oracle[illegible]

(Oacle Warehouse Builder,OWB)[illegible]ETL[illegible]

Python[illegible]

Facebook[illegible]MySQL[illegible]Oracle[illegible]ETL[illegible]

[illegible](T)[illegible]

[illegible]Oracle[illegible]400GB[illegible]

[illegible]

□□□□□

□□□□□□□□□(Iternet Data Center,IDC)□□□□□□□□2011□□□□1800exabyte□1exabyte=2^{16}byte)□□

□□□

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□(WW)□□□□□□□□□□□□□□□□□□□□□□□□Autonomy□□□□□1996□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□(Byesian inference algorithms)□□□□□□□□□□□Fast Search and Transfer(FAST)□□□1997□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□Endeca□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□“□□□□”(fcted search)□□□G□□□□□□□□□□□□□□□□□□2000□□□□□□□□□□□

□□□Endeca□□

□□□1944□□MIT□□Kurt Lewin□□□“□□□□”(ation research)□□□□□□□□

[illegible]Peter Senge[illegible]“[illegible]”(Larning Organization)[illegible]The fifth Discipline[illegible](Boadway Business[illegible]

[illegible]Facebook[illegible]

“□□”□“□□”[1]

□□□□□□□Facebook□□□□□□□□□□□□□□□□□□□□□400GB□□□□□□□□□□□□□□□□□□□□□Oracle□□□□□□□□□□□□□□□□□□□□□□□□□□□□□24□□□(smmary information)□□□□□□

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□Web□□□□□□□□□□□□□□□□□□□□□□□□□□□Cheetah□□

□□□Cheetah□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□(achival storage)□□□□□□□□□□□□□Cheetah□□□□□□□NetApp□□□□□□□□□□□□□□NetApp□□□□□□□□□□□□□□□□□□“□□”(shema)□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□Unix□□□cron□□□Cheetah□□□□□□□□□□□□□□□□□□□□□□□□□Cheetah□□Cheetah□□□

Apache□Hadoop□□□□Doug Cutting□Mike Cafarella□2005□□□□□□□□□□Cheetah□□□□□□□Doug□□□□□□□□□□□Hadoop□□□□□□□□□□Apache 2.0□□□G□□□□□□□□□□□MapReduce□□□□□□2006□1□□□□Doug Cutting□□□□□□□□□□□□□□□Hadoop□□2006□4□□□□□□□188□□□□□□□□47□□□□□1.9T□□□□□□□□□□Hadoop□□□□□□□□□□Cheetah□□□□□□□□□□□□□□□□□□□□□□□2008□4□□Hadoop□910□

□□□□□□209□□1T□□□□□□□□Hadoop□□□□□□□□□□□□□60□Web□□□3□500GB□SATA□□□□□Facebook□□□□Hadoop□□

□□□□□□□□□□Hadoop□Cheetah□□□□□□□□Hadoop□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□Facebook□□Cheetah□□□□□Hadoop□□□□□□□□□□□□□□□□□□□□□Hadoop□

□□Hadoop,Facebook□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□TB□□

□Facebook□□□Hadoop□□□□□□□□□□□□□□□□□Hadoop□□□□□MapReduce□□□□□Facebook□□□□□□□□□□□□□□□□□

Facebook□□□□□□□□□□□(smi-public)□□□□□□□□□□□□□□□□□□□□10□□□□□□□□□□□□□□□□□□□□□□2007□□□□□□□□□□□□□□□□□□Roddy Lindsay□□□□□□□Hadoop,Roddy□□□□□□□□□□□□□□□□□□□□Lexicon□□□□□□□TB□□□□□□□□□□URL:http□//facebook.com/lexicon□□□□

□□Facebook□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□2007□5□□□Facebook□□□□□□□□□□“□□”□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□spam□□□□□□□□□API

[illegible]

[1][illegible](ceetah)[illegible]Facebook[illegible]The Cheetah[illegible](eephant)[illegible]Hadoop[illegible]

[illegible]

[illegible]G[illegible]“[illegible](eaborate)[illegible]”[illegible]G[illegible]

[illegible]Facebook,Hadoop[illegible]Facebook[illegible]Cameron Marlow[illegible]Wikipedia[illegible](caracter trigram frequency counts)[illegible]Facebook[illegible]G[illegible]14[illegible]Peter Norvig[illegible]

G[illegible]“[illegible]”[illegible]

□□□□□□□□

□Facebook□□□□□Hadoop□□□Hadoop□□□□□□□□□□□□□□□□Hive□Hive□□□□□□□□SQL□□□□□MapReduce□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□Hadoop□□Hive□□□□□□□□□□□□□□□□□□□(uer interface,UI)□□□Hive□□□□□Hipal□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□TB□□□□□□□□□□□□□□□□□□□□□□□□Apache 2.0□□□□Hive□□Hadoop□□□□□□□□□□□□□□□□□□□

□□Hive□□□□□□□□□□□□□□□□Argus□□IBM□Many Eyes□□□□□□□□□□□□□Databee□□Python□MapReduce□□□□□PyHive□□□□□□□□□□□□□□□□□□□□Cassandra□□□□□□□□□Apache□□□□□□

□□□□□□□□□□□□□□□□□□□□□Hadoop□□□□□□□□□(mliple tiers)□□□□□□□□□□□□□□□□□□□□□□□□□□Web□□□□□□□□□□□□□□□□□Hadoop□□□□□□□□(HFS)□□□□□□□□□Databee□□□□□□□□□□□□□□□□□□□□□□□Hive□□□HDFS□□□□□□□Hive□□□□□□□□□□□□□□□Argus□□□□□□□□□HDFS□□□□□□home□□□□□“□□”(sndboxes)□□□□□□□□□□(pototype)□□□

□□Hadoop□□□□□25□□□2.5PB)□□□□□□□□□15TB□□□□□□□□□□□3000□□□□MapReduce□□□□□□□□55TB□□□□□□□□□□□□□□□□□□□

□□□a/b□□□□□□□□□□□□□□□□□□□□□Facebook□Hadoop□□□□□□□□□□□□□□□□□□

□□□□□□Itamar Rosenn□Cameron Marlow□□□□□□□□□□□(Ingitudinal study)□□□(last-angle regression)□□□□□□□□□□□□□□□□□Hadoop□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

□□□□□□□□□□□□□□□□□□□□□□□□2009□CHI□□□□□“Feed Me:Motivating Newcomer Contribution in Social Network Sites”□□□□Facebook□□□□□□□□□□□□□□□□□□□□□Facebook□□□□□□□□□□□□□□□“Gesundheit!Modeling Contagion through Facebook News Feed”□□□2009 ICWSM□□□□□

□Facebook□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□Facebook□□□□□□□□□□□□□□□□□□□

MAD□□□Cosmos

2009□VLDB□□□“MAD Skills:New Analysis Practices for Big Data”□□□□□□□Fox Interactive Media(FIM)□□□□□□□□□□Hadoop□Greenplum□□□□□□□□□□FIM□□□□□□□□□□□□□□□□□□□□□□□□□□Facebook□□□□□□□□

□□□□□□□□□FIM□□□□□□□□□□(Mgnetic)□□□(Aile)□□□□(Dep)□“□□”□□“□□”□□“□□□”□□□□□□□□□□□□□□□□□□□

□FIM□□□□□□Greenplum□□□□□□□□□□□(saging)□□□□□□□□□□□□□□□□□□□□□□Facebook□Hadoop□□□□□□□□□□

□□□□□□□□□□□□□(dta management stack)□□□□□□□□“Dryad:Distributed Sata-Parallel Programs from Sequential Building Blocks”□“SCOPE:Easy and Efficient Parallel Processing of Massive Data Sets”□□□□□□□□□□□□□□□Facebook□□□□□□□□□□□□□□□□□□□□□□□□□□□□Cosmos□□□□□□□□□Dryad□□□□□□□□SQL□□□□□SCOPE□

□□API□□□□□□□□□□□□□□□□□□□□□□□□(dtaspace)□□□□□□□

□□□□□□□□□□□□

□□□□Quantcast□Last.fm□□□□□□□□□□□□□□(ptabyte-scale)□□□API□□□□□□□□□□□□□□□□□□□□□□□□□□□□

□□□

□□□“□□□□”□□□“From Databases to Dataspaces:A New Abstraction for Information Management”(htp□//www.eecs.berkeley.edu/□franklin/Papers/dataspaceSR.pdf)□□□□□□□□□□□□□□□□□□□□□□□□□□□□□API□□□□□□□□API□□□□□□□□□□□□□□□□□□□

□□□API□□

□□(te Data Scientist)□□□□

[illegible]

[illegible]G[illegible]Hal Varian[illegible]

[illegible]Varian[illegible]“[illegible]

[illegible]”

[illegible]Facebook[illegible]

[illegible]Python[illegible]

[illegible]R[illegible]Hadoop[illegible]

[illegible]

[illegible]“[illegible]”[illegible]

[illegible]

(Qants)[illegible]

Facebook[illegible]

[illegible]CERN[illegible]Large Hadron Collider[illegible]

[illegible]

[illegible]Davenport[illegible]Harris[illegible]Competing on Analytics[illegible]

[illegible]2007[illegible]Baker[illegible]The Numerati[illegible](Hughton Mifflin

Harcourt[illegible]2008[illegible]Ayres[illegible]Super Crunchers[illegible](Bntam[illegible]2008[illegible]

[illegible]

[illegible]

[illegible]

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□“□□□”□□□□□□□□□□□□□
□□□□

[illegible]

[illegible]Facebook[illegible]

[illegible]“[illegible]”[illegible](hman component)

[illegible]

[illegible]

[illegible]6[illegible] [illegible]

Jason Dykes[illegible]JoWood

[illegible]

[illegible]“[illegible]”[illegible](pace)[illegible](lcation)[illegible]80%[illegible](McEachren[illegible]Kraak 2001[illegible]

[illegible]G[illegible]“[illegible]”(Snsemaking)[illegible](Adrienko[illegible]2008[illegible]100[illegible]GPS[illegible]iPhone[illegible]

[illegible]“[illegible]”[illegible]

[illegible](Klawa 2007[illegible]

□□□□□Geograph□□

□□□□Geograph□□[1]□□□□□□□□□□□□□□□□□□□□(Bitish Isles)□□□

Geograph□□□□Gary Rogers□□□□Paul Dixon□Barry Hunter□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□“□□□□□□□□□”(Gograph 2009□□□□□□□□□□□□□8500□□□□□□□□□120□□□□□□□□244000□□□□□□□90%□□□6-1□□□□□□□□□□□□□□□□□——□□□□□□□□□□□□“□□□”□



□ 6-1 □Aberuchill□□□□□□□□□Bioran Dalchonzie□□□Geograph □□□□□□□□□ 100 □□□□

(htp □ //www.geograph.org.uk/photo/1006884 □ □ □ □ □ □ □ Dr.Richard Murray □ □ □ □ Creative Commons License(http□//creativecommons.org/licenses/by-sa/2.0/□□□□□□□□□12□

Geograph□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□1km□□□□□□□□□□□□□□□□□□□□

□□(Hwgood□2007□□

□□□□□□□□□□□

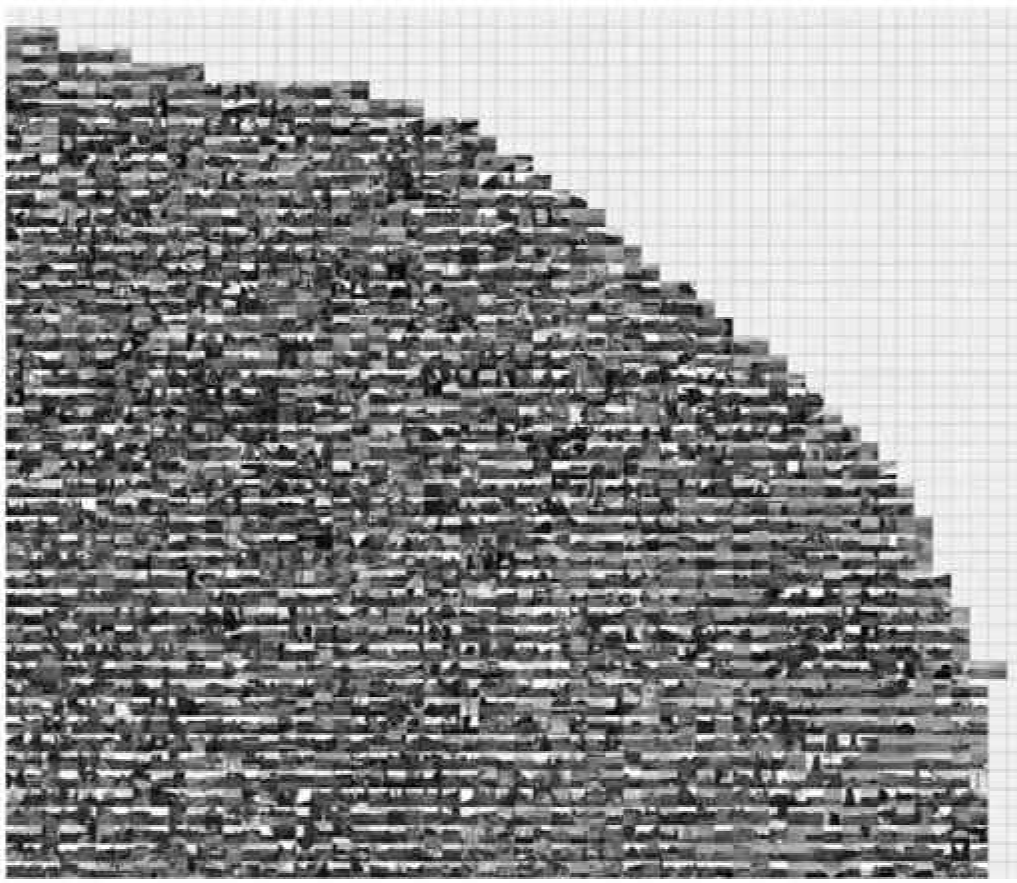
□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□·□□□□□□□□□□□□□□□□□□□□□□□□“□□”□□

·□□“□□□□”(bttom-up)□“□□□□”(cllective effort)□□□“□□□□□□□□□□□”(ctizens as sensors)(Godchild 2007□□

·□□□□□□□□□□□□□□□□□Web□□□

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□·□□□□□□□□□□□□□□□6-2□□□

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□·□□□□□□□□□□□□□□□□□□□□Geograph□□□□□□□□“□□□”□□□“□□□□”(te geograph of the year,GOTY)□□□□□□□□□□□□□□□□□□□□



□ 6-2□Norfolk□□□□□□□□□□□□□□□□□□□□□□□□□□□□□1□□□□ □ □ □ □ □ □ □ Creative Commons License(http□//creativecommons.org/licenses/by-sa/2.0□□□□□□□□□13□

Geograph□□□Geograph□□□□□□□□□□□□□□□□□Geograph□□□□

[illegible]Geograph[illegible]

[illegible]

[1] [illegible] Web [illegible] http[illegible]//www.geograph.org.uk/[illegible]

[illegible]

[illegible]Geograph[illegible]15[illegible]——[illegible](Fsher 1998[illegible]Tufte[illegible]“[illegible]”(Tfte 1983[illegible]

[illegible]

[illegible]Kosara 2007[illegible]Beautiful Code[illegible](Oam[illegible]Wilson 2007[illegible]Brian Kernighan(Kernighan 2007[illegible]“[illegible]”[illegible]Yukihiro Matsumoto[illegible]Ruby[illegible](Mtsumoto 2007[illegible]“[illegible]”[illegible]

□□□□□□□□□□□□□□□□□——□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□
□□Kernighan□□□□□□□□□□□□□“□□”□□□□□□□□□□□□□□□□□□□□□□□
□□□
□□
□□□□□□□□□□(Gstalt)□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□
□□□

Tufte□1983□□□□“□□/□□”□□□□□——□□□□□□□□□□□□□□□□□□□□
□□
□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□/□□□□□□□□□□□□□□□□□
□□
□□□
□□
□□
□□□□□□□□□□□□□□□□□□□□□□□□——□□□□□□□□□□

□□□□□□□□□□□□□□

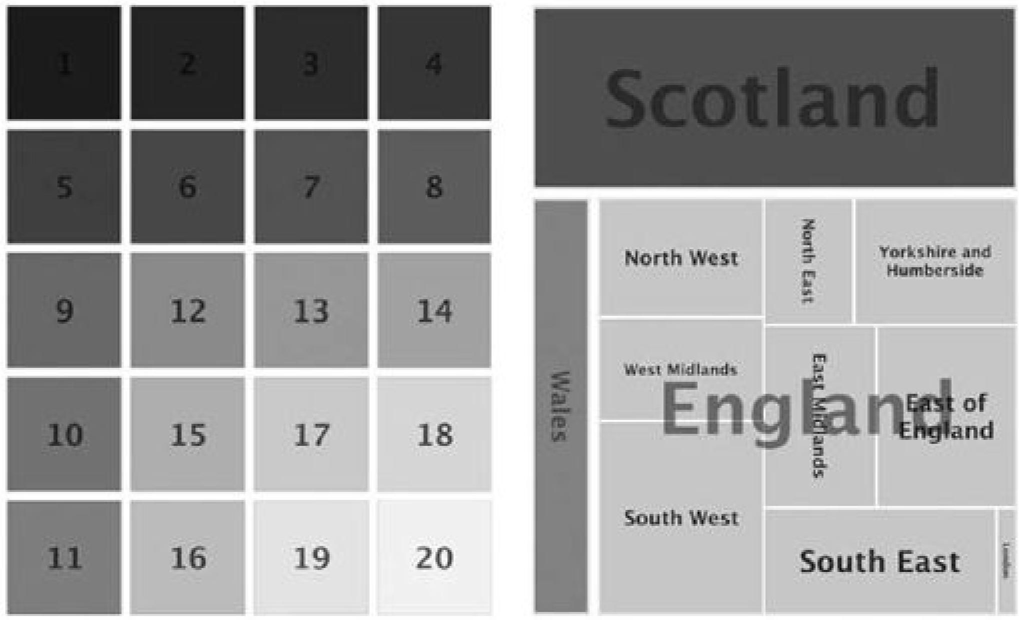
□□□□□□□□□□□□(sace-filling)□□□□(Sneiderman 1992□□□□6-3□□
□□
□□□□□□□“□”□□□□□□□□□□□□□□□□□□□□□□□□□□“□”□□□□□□□□□□□□
□□
□□

[illegible]

[illegible]Geograph[illegible]

[illegible](Cwthon[illegible]vande Moere 2007[illegible]

[illegible]“[illegible]/[illegible]”[illegible]“[illegible]/[illegible]”[illegible]6-3[illegible](Brtin 1983[illegible]“[illegible]”(Fbrikant[illegible]2002[illegible]



[illegible] 6-3[illegible](suarified)[illegible]20[illegible]1[illegible]20[illegible]

[illegible](Wod[illegible]Dykes 2008[illegible]

Geograph[illegible]

“[illegible]”(pace)[illegible] Geograph[illegible]

Ross Purves[illegible]Alistair Edwardes[illegible]Zurich[illegible] Geograph[illegible]“[illegible]”[illegible]“[illegible]”[illegible]“[illegible]”[illegible]Geograph[illegible](Ewardes[illegible]Purves 2007[illegible]

[illegible]Edwardes[illegible]Purves[illegible]Geograph[illegible](Ewardes[illegible]Purves 2007[illegible]

[illegible]Geograph[illegible]Geograph[illegible]

[illegible]Geograph[illegible]

[illegible]

[illegible]

Geograph[illegible]2008[illegible]4[illegible]75[illegible]

[illegible]Edwardes[illegible]Purves[illegible]

[illegible]

Geograph[illegible]6-4[illegible]

[illegible](Wod[illegible]Dykes 2008[illegible]



[illegible] 6-4[illegible]

[illegible]/[illegible](fcet)/[illegible]“[illegible]”[illegible]

[illegible]14[illegible]

[illegible]6-4[illegible]

[illegible]

[illegible]

[illegible](hll)[illegible]

[illegible]

[illegible]“[illegible]”[illegible]“[illegible]”[illegible]“[illegible]”[illegible]

[illegible]

[illegible]

[illegible]1000[illegible]

[illegible]Geograph[illegible]

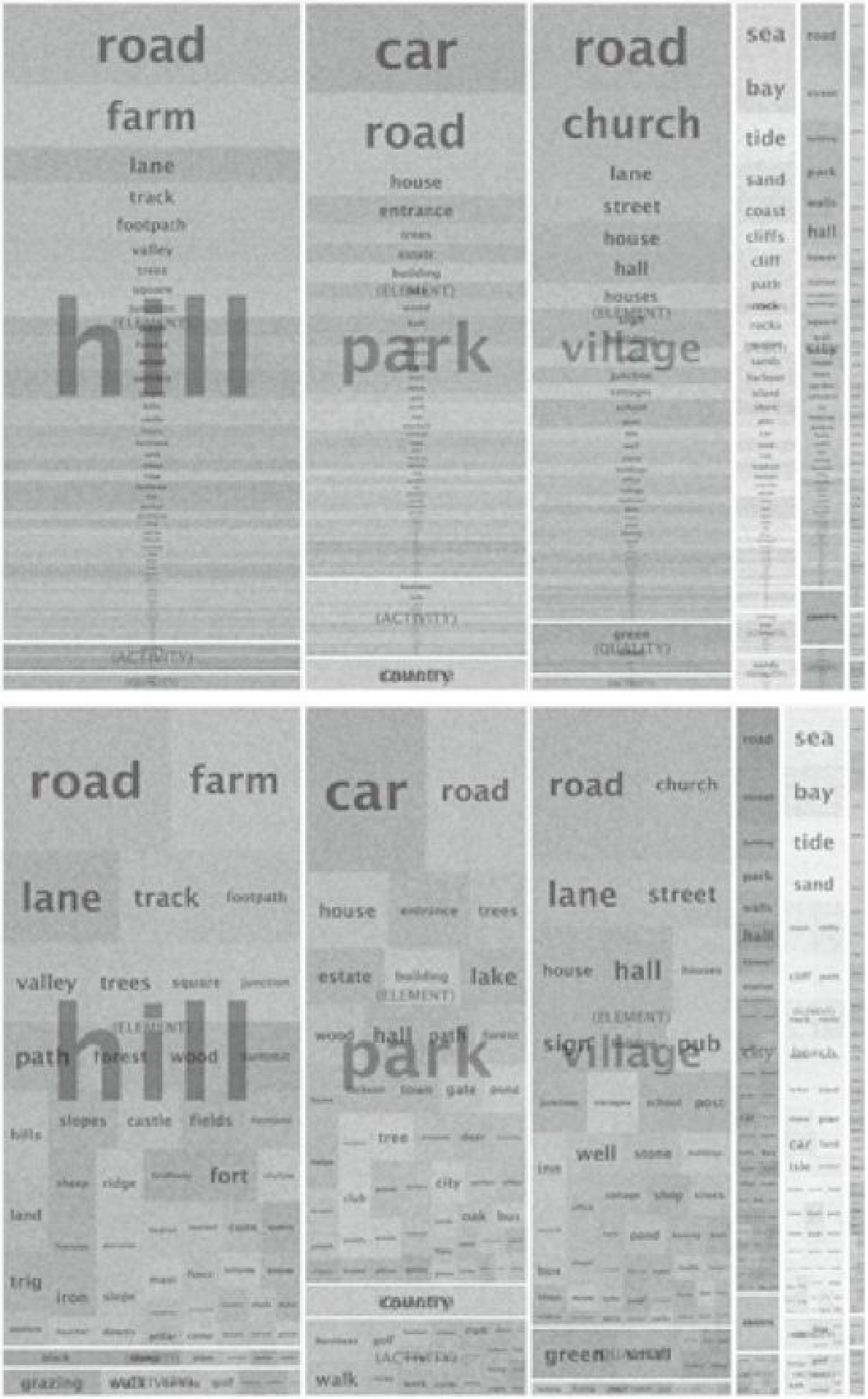
[illegible]“[illegible]”[illegible]

[illegible]6-5[illegible]

[illegible]

[illegible]6-5[illegible]

[illegible]“[illegible]”[illegible]


road
farm
lane
track
footpath
valley
trees
hill
(ACTIVITY)
car
road
house
entrance
park
(ACTIVITY)
country
road
church
lane
street
house
hall
houses
village
green
(QUALITY)
sea
bay
tide
sand
coast
cliffs
cliff
hall
road
farm
lane
track
footpath
valley
trees
square
junction
(ELEMENT)
hill
path
forest
wood
summit
hills
slopes
castle
fields
sheep
ridge
fort
land
trig
iron
grazing
car
road
house
entrance
trees
estate
building
lake
(ELEMENT)
wood
hall
path
park
town
gate
tree
club
city
oak
country
walk
golf
road
church
lane
street
house
hall
houses
(ELEMENT)
sign
village
pub
junction
school
post
inn
well
stone
shop
pond
box
green
sea
bay
tide
sand
cliff
path
car
isle

[illegible] 6-5[illegible]/[illegible]/[illegible]“[illegible]”[illegible]“[illegible]/[illegible]”[illegible]15[illegible]

[illegible]

[illegible]6-4[illegible]6-5[illegible]

[illegible]Geograph[illegible]“[illegible]”(esting)[illegible]“[illegible]”(nrthing)[illegible](Bitish National Grid)[illegible]

[illegible]CIELab[illegible]a[illegible]b[illegible]“[illegible]”[illegible]“[illegible]”[illegible]6-6[illegible]6-4[illegible]



[illegible] 6-6[illegible]“[illegible]”[illegible]CIELab[illegible]

[illegible]16[illegible]

[illegible]

Geograph[illegible]

[illegible]

[illegible]

[illegible]

[illegible]-[illegible](lcation-place)[illegible]

[illegible]

[illegible]6-6[illegible]

[illegible]

[illegible]“[illegible]/[illegible]”[illegible]
CIELab[illegible]6-7[illegible]

[illegible]

[illegible]6-8[illegible]

[illegible]

[illegible]CIELab[illegible]6-7[illegible]/[illegible]/[illegible]/[illegible]/[illegible]Skupin[illegible]Fabrikant[illegible]2003[illegible]



□ 6-7□“□□□□□□”□□□□□□CIELab□□17□

□6-8□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□-□□6-8□□□6-9□□6-9□□□□□□□□□□□□□□□□□□□□□□□□□□□□

□□□

[illegible]

[illegible]

[illegible]6-10[illegible]

[illegible]

[illegible]/[illegible]/[illegible]/[illegible]/[illegible]



[illegible] 6-8[illegible]

[illegible]CIELab[illegible]

[illegible]18[illegible]

□□□□

□□□□□□□□□□□□□□□□□□□□□□□□□□□“□□/□□□□□□”□□□□□

□□□

□□□□□□□□□□□□Matsumoto□2007□□□□□□□□□□□□□□□□□□□

□□□

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

Geograph□□□□□□□□□□□□□□□□Geograph□□□□□□□

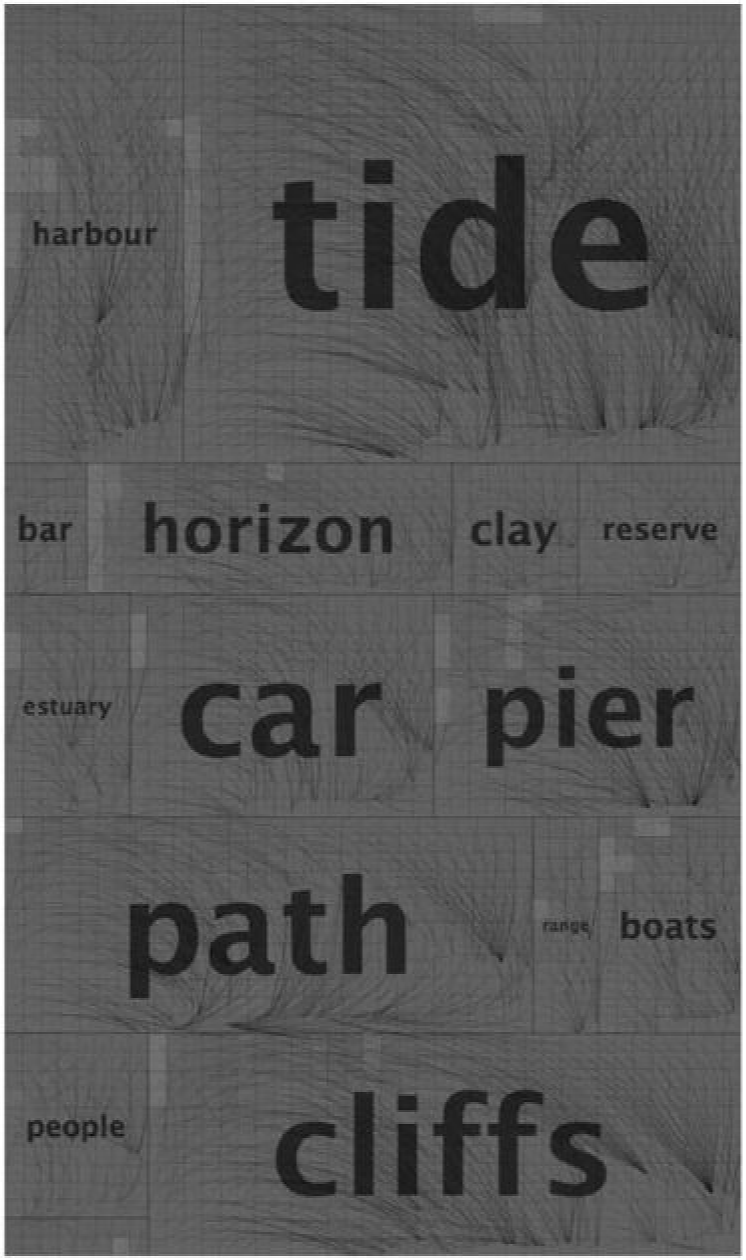
Geograph□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

□□



[illegible] 6-9[illegible]19[illegible]

[illegible]Geograph[illegible]Geograph[illegible]——[illegible]“[illegible]”[illegible]



[illegible] 6-10[illegible]beach[illegible]

[illegible]6-9[illegible]20[illegible]

[illegible]

[illegible]

[illegible]Geograph[illegible]

[illegible]
[illegible]
[illegible]

[illegible]

[illegible]

[illegible]

[illegible]Geograph[illegible]

[illegible]Geograph[illegible]“[illegible]/[illegible]”[illegible]Geograph[illegible]

□□

□□□Barry Hunter□Geograph British Isles(http□//www.geograph.org.uk/credits/2008-04-31□□□□□□□□□□□□□□□□□□□□□Creative Commons Attribution-ShareAlike 2.5 License(http□//creativecommons.org/licenses/by-sa/2.5/□□□□□□

□□□Ross Purves□Ross Purves□Geograph□□□□□□□□□□□□□□□□□□□TRIPOD□□□□□□□□□□□□□□□□□□□□045335□

□□□□

[1]AndrienkO'G.□N. AndrienkO'J.Dykes,S.Fabrikant,and M.Wachowicz.2008."Geovisualization of dynamics,movement and change:key issues and developing approaches in visualization research."Information Visualization,v.7□173-180.

[2]Bertin,J. 1983.Semiology of graphics(W.J.Berg,trans.□.Madison:University of Wisconsin Press.(Oiginal work published 1973.□

[3]Cawthon,N.□and A. vande Moere.2007."The Effect of Aesthetic on the Usability of Data Visualization□"Proceedings of the 11th International Conference Information Visualization,IEEE Computer Society Washington,DC□637-648.

[4]Edwardes,A.□and R. Purves.2007."A theoretical grounding for semantic descriptions of place."M.Ware and G.Taylor(eds.□.LNCS:Proceedings of 7th Intl,Workshop on Web and Wireless GIS,W2GIS□106-120.

[5]Fabrikant,S.□M. RuoccO'R.Middleton,D.MontellO'and C.Jörgensen.2002."The first law of cognitive

geography:Distance and similarity in semantic space."Proceedings of GIScience 2002，31-33.

[6]Fisher,P. F.1998."Is GIS hidebound by the legacy of cartography?"The Cartographic Journal,v.35，5-9.

[7]Geograph."Geograph British Isles-photograph every grid square."http：//www. geograph.co.uk/(acessed April 9，2009）.

[8]Goodchild,M. 2007."Citizens as sensors:the world of volunteered geography."GeoJournal,v.69，211-221.

[9]Hawgood,D.，D. Dunford,R.Farrow,B.Hunter,and P.Mayes."Geograph or supplemental."http：//www.geograph.org.uk/article/Geograph-or-supplemental/(acessed April 9，2009）.

[10]Kernighan,B. W.2007."A Regular Expression Matcher，"in Beautiful Code:Leading Programmers Explain How They Think,ed.Andy Oram and Greg Wilson，1-9.Sebastopol,CA:O'Reilly.

[11]Kosara,R. 2007"Visualization Criticism-The Missing Link Between Information Visualization and Art."Proceedings of the 11th International Conference Information

Visualization,IEEE Computer Society Washington,DC□631-636.

[12]MacEachren,A. M.□and M.J.Kraak.2001."Research challenges in geovisualization."Cartography and Geographic Information Science,v.28□3-12.

[13]MatsumotO'Y. 2007."Treating Code as an Essay□"in Beautiful Code:Leading Programmers Explain How They Think,ed.Andy Oram and Greg Wilson□477-481.Sebastopol,CA:O'Reilly.

[14]Oram,A. and G.Wilson,eds.2007.Beautiful Code:Leading Programmers Explain How They Think.Sebastopol,CA:O'Reilly.

[15]Shneiderman,B. 1992."Tree visualization with tree-maps□2-d space-filling approach."ACM Transactions on Graphics(TOG)□v.11□92-99.

[16]Skupin,A. and S.Fabrikant.2003."Spatialization Methods:A Cartographic Research Agenda for Non-geographic Information Visualization."Cartography and Geographic Information Science,v.30□99-119.

[17]Tufte,E. R.1983.The Visual Display of Quantitative Information(First Edition).Cheshire,CT:Graphics Press.

[18]Wood,J. and J.Dykes.2008."Spatially Ordered Treemaps."IEEE Transactions on Visualization and Computer Graphics,v.14：1348-1355.

□7□　□□□□□

Jeff Jonas□Lisa Sokol

[illegible]

[illegible]“[illegible]”[illegible]

[illegible]——[illegible]50[illegible]

[illegible]“[illegible]”[illegible]

[illegible]——[illegible]

[illegible]“[illegible]”[illegible]

[illegible]

[illegible]“[illegible]”[illegible]“[illegible]”[illegible]

[illegible]

[illegible]7[illegible]“[illegible]”[illegible]2[illegible]

[illegible]

[illegible]Amazon[illegible]

[illegible]

[illegible]Web[illegible]Web[illegible]“[illegible]”[illegible]

[illegible]“[illegible]”[illegible]

[illegible]“[illegible]”[illegible]“[illegible]”[illegible]“[illegible]”[illegible]“[illegible]”[illegible]

[illegible]

[illegible]——[illegible]

[illegible]20[illegible]90[illegible](rverboat gaming)[illegible]

[illegible]Bossier[illegible]

[illegible]——[illegible]

[illegible]

[illegible]

[illegible]20[illegible]30[illegible]40[illegible](Cime Scene Investigation,CSI)[illegible]

[illegible](Tipwire)[illegible][1][illegible]——[illegible]——[illegible]

[illegible](rulette table)[illegible]“[illegible]”(pst posting)——[illegible]

[illegible]“[illegible]”[illegible]32[illegible]“[illegible]”[illegible]

[illegible]“[illegible]”[illegible]——[illegible]“[illegible]”[illegible]“[illegible]”[illegible]——[illegible]

[illegible]

[illegible]“[illegible]”[illegible]

[illegible]

[illegible][2][illegible]“5764[illegible]44-00321[illegible]”

[illegible]“[illegible]”[illegible]“[illegible]”[illegible]

[illegible]“[illegible]”[illegible]——[illegible]90[illegible]

[illegible]

[illegible]

[illegible]

[illegible]“[illegible]”(nnobvious relationship)[illegible]

[illegible]“[illegible]”[illegible]

[1][illegible]

[2] [illegible]“[illegible]”(Nn-Obvious Relationship Awareness,NORA)[illegible]Systems Research[illegible]Development[illegible]SRD[illegible]Jeff Jonas[illegible]SRD[illegible]IBM[illegible]IBM[illegible]——IEEE[illegible]Threat[illegible]Fraud Intelligence-Las Vegas Style(http[illegible]//jeffjonas.typepad.com/jeff_jonas/2006/11/ieee_paper_thre.html)[illegible]

[illegible]

[illegible]

[illegible]“[illegible]”[illegible]

[illegible]

[illegible]

[illegible]

[illegible]“[illegible]”[illegible]

[illegible]

[illegible]·[illegible]ID[illegible]

[illegible]

“[illegible]”[illegible]——[illegible]

[illegible]·[illegible]——[illegible]

[illegible]2000[illegible]8[illegible]

[illegible]8[illegible]

[illegible]8[illegible]

[illegible]“[illegible]”[illegible]

[illegible]

[illegible]

[illegible]eBay[illegible]Dewey[illegible]“[illegible]”[illegible]

G[illegible]G[illegible]

[illegible]“[illegible]”[illegible]

[illegible]“[illegible]”[illegible]

[illegible]“Billy the Kid”[illegible]

[illegible][1][illegible]“Billy the Kid”[illegible]“[illegible]”[illegible]“William Antrim”[illegible]“[illegible]”[illegible]

[illegible]Email[illegible]Email[illegible]——[illegible]Email[illegible][2][illegible]

[illegible]“[illegible]”(rlationship-aware)[illegible]Billy theKid[illegible]William Antrim)[illegible]William Antrim[illegible]Billy the Kid[illegible](bndle)[illegible]Kevin Bacon[illegible]

[illegible]“[illegible]”[illegible]

[illegible](Prsistent context)[illegible]“[illegible]”[illegible]

[illegible]“[illegible]”[illegible](Jst-in-time context)[illegible]

[illegible]“[illegible]”[illegible]

[illegible]4 exabytes[illegible]1 exabytes=2^{16}[illegible]5TB[illegible]

[1][illegible]“[illegible]”[illegible]

[2][illegible]/[illegible]/[illegible]“Entity Resolution Systems vs.Match Merge/Merge Purge/List De-duplication Systems”(htp[illegible]//jeffjonas.typepad.com/jeff_jonas/2007/09/entity-resoluti.html)[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]“[illegible]”[illegible]8[illegible]

·[illegible]

·[illegible]

·[illegible]

·[illegible]

·[illegible]

·[illegible]

·[illegible]/[illegible]

·[illegible]

[illegible]

[illegible]“[illegible]”[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible](cntextualized)[illegible]1[illegible]2[illegible]“[illegible]”[illegible]3[illegible]

[illegible]

[illegible]

[illegible]Bob Joneses[illegible]

[illegible]

[illegible]

[illegible]——[illegible]——[illegible]

[illegible]/[illegible]

[illegible]“[illegible]”[illegible]“[illegible]”[illegible]

[illegible]1[illegible]

[illegible]/[illegible]

[illegible]

[illegible]50[illegible]

□□□□

□□□□□□□□□□□“□□□□□□□”□□

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

·□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

·□□[1]□

·□□□

·□□□□□“□□□”(aonymized)□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□[2]□

[1]□□□□□□□□□□□□□□□□□□□□□□□□Markle Foundation□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□http□//www.markle.org/downloadable_assets/nstf_IAL_020906.pdf□

[2]□□□□“□□□”□□□□□□□Jeff Jonas□John Karat□□□□Emergent Information Technologies and Enabling Policies for Counter-

Terrorism□□□□Robert L.Popp□John Yen□2006□Wiley-IEEE Press□□□□“Anonymized Semantic Directories.A Privacy-Enhancing Architecture for Enterprise Discovery”□□□□□(htp□//www.wiley.com/WileyCDA/WileyTitle/productCd-0471776157.html)□

[illegible]

“[illegible]”[illegible]/[illegible]

[illegible]

[illegible]“[illegible]”(cgnitive computing)[illegible]

[illegible]8[illegible] [illegible]

JudValeski

□□

□□□□□□□□□□□□“□□”□□□□□□□□□□□□□□□□“□□”□□□□□□□CAT-5□□□□□IP□□□□□DNS□Sockets□□-□□□□□□□□□□□□□□“□□”□□□□□□□□□□□□□□□□□□HTTP□□□□□□□□□□□□□□□□“□□”□□□□

□□□□□15□□20□□90□□□□□□□□□□□“Web□□□”□□□□□□□□□□□□□□□□□□□□□□□□□□□——□□□2009□1□14□□PEW□□□Adults and Social Network Websites□□□□18□34□□□□□□□□50%□□□□□□□□□□□□□□[1]□

□□□“□□□□□”(Ocam's razor)□□□□□□□□□□□□□□□□□□□□□180□□HTTP□□□□□POST□□□“□□”□□□□□□□□□□□□□□□□□□□□□□Gnip(http□//www.gnip.com)□□□□□□□□□□□□□□□□□□□□□□□

[1]http □ //www.pewinternet.org/Reports/2009/Adults-and-Social-Network-Websites.aspx

Socket

HTTP HTTP

HTTP

HTTP

HTTP

iPhone Android HTTP (Nkia

S60 HTTP HTTP

(XPP)

AOL

P2P XMPP

XMPP

IM PubSub 2008 Cisco Jabber.com(XMPP

XMPP

BitTorrent

□□□□□□□□□□□□□□□□□□“□□”□□□□□□□□□□□□□□
□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□
BitTorrent□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□
□□□□□□□□□□

□□/P2P

HTTP□□□□□□□□□□□/□□□□□□□□□□□□□□□□□□□□
□□□□□□□□□n□□□□□□□□□□□□□□□□□□□□□□□□□□
□□□□□□□□□□□□□□□□□□□□□□□□□□□“□□□”(Pint-2-Point□Peer-2-Peer)□□□□□□□□□□□□□□□□

□HTTP□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□
□□□□□□□□□□□□□□□□□□□□□□“□□”□

□

XML□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□
□□□□XML□□□□□□□□□□□□□□□□□□□□□□□□□□XML
□□□□□□□□□□□□□□□□□□□□XML□□□□□□□□□□□□□□
□□□□□□□□□□□□□□□□□□□□□□□□□□□AOL□□FDO□
□FDO□□□□□□□□□□□□/□□□□300□□□□□□□□□□□□□
□□□□□□□□□□□□□□□□

□□□□□HTTP□□□□XML□□□□□□□□□□□□15□□□□□□
□□□□□□□□□□□——□□□□□□□□□□□□□□□□□□□□□□□
□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

[illegible]“[illegible]”[illegible]“[illegible]”[illegible]

[illegible]XML)[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible](AI)[illegible]HTTP[illegible]API[illegible]

[illegible]REST[illegible]Web[illegible]Roy Fielding[illegible]HTTP1.1

[illegible](Rpresentational State

Transfer,REST)[illegible]

[illegible]CORBA[illegible]DCOM[illegible]

[illegible]HTTP[illegible]Web[illegible]

[illegible]We b[illegible]API[illegible]8-1[illegible]API[illegible]

[illegible]

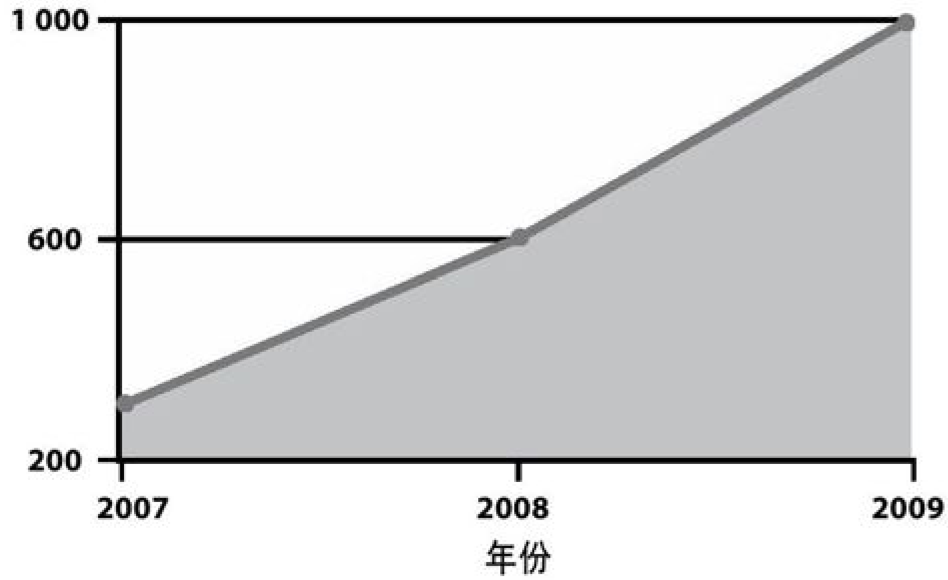
[illegible]Programmable Web(http[illegible]//www.programmableweb.com/[illegible]

[illegible]API[illegible]

[illegible]API[illegible]WAN[illegible]

[illegible]API[illegible]Web API[illegible](Srvice-Level

Agreement,SLA)[illegible]API[illegible]



□　8-1□□□□□□□□□API□□□□□□□□□□□□Programmable Web□□

□□□□□□□□□□□□□□□□□API□□□□□□□□□□□□□□□□□□□□□□□□□□□□
□□□□□□

□□□□□□□API□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□
SLA□□“□□□
□”(mshups)□□□□□□□□□□□□□□“□□”□□□□□□□□□□□□□□□□□□□□□□□
□□□
□□□□□□□□□□□□□□□□□□□□□□□□□□□□“Comet”□“Web Hooks”□□□
□□“PubSub”□□
□□□□□□□□□□“□□”□□□□□□□□□□□□□□□□“□□”□□□□□□□□

□□□□□□□□□□□□□□□□□(tme-flow-processing)□□(Pimitives)□□
□(Plling)□□□(Eents)□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□
□□□

[illegible]

[illegible]

[illegible]1[illegible]10[illegible]“[illegible]”“[illegible]”“[illegible]”“[illegible]”[illegible]“[illegible]”[illegible]

[illegible][1]

[illegible]API[illegible]API[illegible]API[illegible]“Web API”[illegible]“Web[illegible]”[illegible]

[illegible]API[illegible]IP[illegible]API[illegible]API[illegible]“[illegible]”(Trottling Layer)[illegible]“[illegible]”[illegible]IP[illegible]“[illegible]”[illegible]API[illegible]Digg)[illegible]

[illegible]

[illegible]“[illegible]”[illegible]90%[illegible]“[illegible]”[illegible]

[illegible]Flickr[illegible]

Flickr[illegible]

[illegible]0[illegible]

[illegible]“Flickr,Friendfeed[illegible]”[illegible](Fickr)[illegible](Fiendfeed)[illegible]“[illegible]”(ofbox)[illegible]

[illegible]2008[illegible]7[illegible]Even Henshaw-Plath[illegible]Kellan Elliot-McCrea[illegible](Prtland)[illegible]OSCON(Open Source Convention,O'Reilly[illegible]“Beyond

REST?”(htp[illegible]//www.slideshare.net/kellan/beyond-rest)[illegible]Friendfeed[illegible]4.6[illegible]24[illegible]

Friendfeed[illegible]Flickr[illegible]API 290[illegible]4.6[illegible]7000[illegible]Flickr[illegible]

[illegible]“1-9-90”[illegible]100%[illegible]1%[illegible]“[illegible]”[illegible]9%[illegible]“[illegible]”[illegible]90%[illegible]“[illegible]”[illegible]Flickr/Friendfeed[illegible]7000[illegible]Flickr[illegible]700[illegible]Friendfeed[illegible]Flickr[illegible]290[illegible]700[illegible]“[illegible]”[illegible]0.02%[illegible]4000[illegible]1[illegible]

[illegible]Flickr/Friendfeed[illegible]4000[illegible]“[illegible]”[illegible]1[illegible]“[illegible]”[illegible]

[illegible]“[illegible]”(O-box)API[illegible]“[illegible]”[illegible]Web API[illegible]

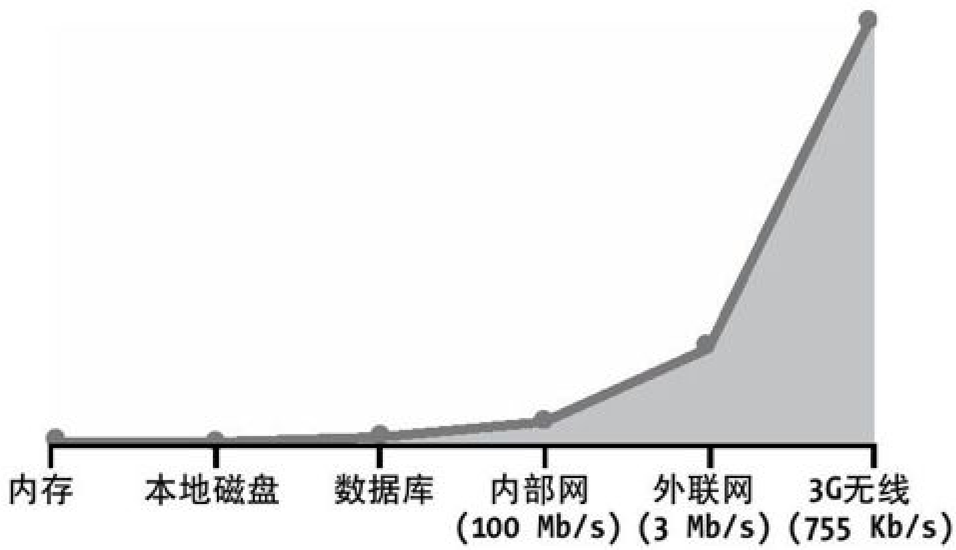
[illegible]“[illegible]”[illegible]

[illegible]

[illegible]“[illegible]”[illegible]

[illegible]

[illegible] API [illegible] I/O [illegible] I/O [illegible] 8-2 [illegible] I/O [illegible] IP [illegible] / [illegible]



[illegible] 8-2 [illegible]

[illegible]

[illegible]4000[illegible]“[illegible]”[illegible]1[illegible]Web[illegible]HTTP/REST[illegible]-[illegible]Web[illegible]Web[illegible]“[illegible]”[illegible]

HTML 5[illegible]

[illegible]HTML5[illegible]“Web Socket”[illegible]XMLHttpRequest[illegible]GUI Web[illegible]“Web Sockets”[illegible]Web[illegible]socket[illegible]

WAN[illegible]

[illegible]“[illegible]”[illegible]Tibco[illegible]Internet[illegible]Web[illegible]

[illegible]HTTP POST[illegible]WAN[illegible](XL)[illegible]Jeff Lindsay[illegible]HTTP POST[illegible]

[illegible]“Web Hooks”(htp[illegible]//blog.webhooks.org/[illegible]HTTP[illegible]XML[illegible]

[illegible](Bta[illegible]VHS)[illegible]“Web[illegible]”[illegible]

Gnip[illegible][2][illegible]Gnip[illegible]

[illegible]API[illegible]Gnip[illegible]API[illegible]API[illegible]Gnip[illegible]

[1][illegible]

[2][illegible]http[illegible]//www.gnip.com/[illegible]

[illegible]

[illegible]HTTP POST[illegible]——Web Hooks——[illegible]API[illegible]XML[illegible]API[illegible]“[illegible]”[illegible]API[illegible]——[illegible]XML[illegible]XML[illegible]bug[illegible]

[illegible]HTML[illegible]Web[illegible]bug[illegible]

[illegible]A[illegible]B[illegible]

[illegible]“[illegible]”[illegible]XML[illegible]XML[illegible]“[illegible]”[illegible]

[illegible]Delicious[illegible]

```
<item>
<title>Fractals derived from Newton-Raphson iteration</title>
<pubDate>Mon, 19 Jan 200920:02:05+0000</pubDate>
<guid isPermaLink="false">http://delicious.com/url/7549fded443f#joe</guid>
<link>http://www.chiark.greenend.org.uk/~sgtatham/newton/</link>
```

```
□dc:creator□iacovibus□/dc:creator□
□comments□http□//delicious.com/url/7549fded443f□/comments□
□wfw:commentRss□http□//feeds.delicious.com/v2/rss/url/a□/wfw:commentRss
□
□source url="http□//feeds.delicious.com/v2/rss/joe"□joe's bookmarks□/source
□
□category domain="http□//delicious.com/joe/"□mathematics□/category□
□category domain="http□//delicious.com/joe/"□newton-raphson□/category□
□category domain="http□//delicious.com/joe/"□fractals□/category□
□category domain="http□//delicious.com/joe/"□iteration□/category□
□/item□
```

□□givealink.org□□□

```
□item□
□title□Bus slams into shop houses after driver collapses behind wheel□/title□
□link□http□//www.thaivisa.com/forum/Bus-Slams-Shop-Coll-t198228.html□/link
□
□description□Bus slams into shop collapses behind wheel□/description□
□/item□
```

□□□□□□□□□□□API□Feeds□□□□□□□□(Uer-Generated Content,UGC)□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□DiSo□□□□□API□□□□□□□□□□□□□□□□□□“□□”□□□□□□□□□□http□//diso-project.org/wiki/activity-streams□

□□□□□□□□□□□□□□□□Gnip□□□Gnip□□“□□”□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

□□□□□□

□□□“□□□□”□□

API

Internet /

API

“freegan”

0.01

API

REST URL

API API

[illegible]HTTP Basic-Auth[illegible]
URL[illegible]“API[illegible]”[illegible]API[illegible]API[illegible]
[illegible]HT TP[illegible]API[illegible]
[illegible]“[illegible]”[illegible]

[illegible]“[illegible]”[illegible]

[illegible]“[illegible]”[illegible]API[illegible]OAuth[illegible]Blind URL[illegible]

OAuth(http[illegible]//oauth. net/[illegible](Tken)[illegible]

Blind URL[illegible]“[illegible]”[illegible]Flickr[illegible]“Guest Pass”[illegible]URL[illegible]Flickr[illegible]“[illegible]”[illegible]Flickr[illegible]“[illegible]”[illegible]“[illegible]”[illegible]Blind URL[illegible]URL[illegible]

Gnip[illegible]

[illegible]

[illegible]Gnip[illegible]

[illegible]

□□□□□□Gnip□□

□□“□□”□□□“□□”□□□□□□API“□□□□”□□□□□API□□□□□□□□□□□□□□□□□□□□□□□□□□Gnip□□□□□□□□□□□□□□□□“□□Web□□”□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□Gnip□□□□□□□□□□□□□□□□□□□□Gnip□□□□□□□□□□□□□□□□□60□□□□□□□□□□□□□□□□□Gnip□□□“□□”□“□□□□”□□□□□□□□□□□□Gnip□□□□□□□□□□Gnip□□□□□□□□□□□□□□□□□□□□□□□□□□□

□□□□□Gnip□□□□□□□□□□□□□□API□□□□“□□”□□Gnip□□□Gnip□□□/□□□□□□□□□□□□□□□□□□□□□□□□□□

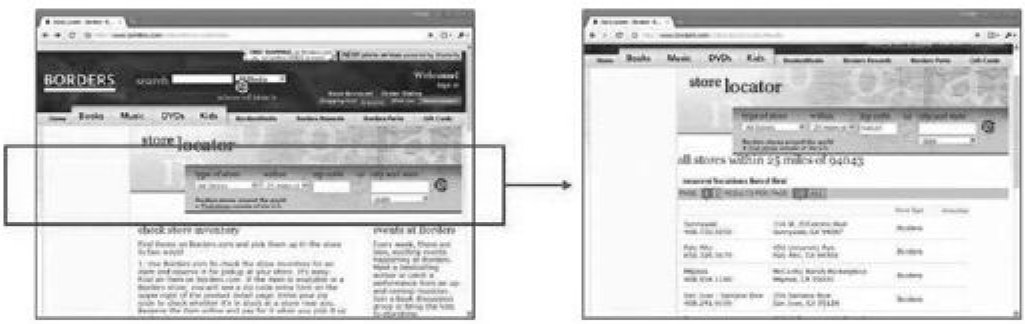
□□□Web□

□9□　□□Deep Web

Alon Halevy□Jayant Madhaven

□□□Deep Web

“Deep Web”□□□□□□□□□□HTML□□□□Web□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□9-1□□□□□□□□□□□□□94043□□□□□□□□□□□□□□□□□□Web□□□□□□□□□□Deep Web□Web□□□□□□□□



□　9-1□Borders□□□□□□□□□□□□□□□□Deep Web□□□□□□□□□21□

Deep Web□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□Web□□(cawler)□□□□□□□□□□□□□□□□□□□□□□□□□□□□□Web□□□□□□□Web□□□□□□□□□□□Web□□□□□□□□□□□□□□□□□□□□□□□□Web□□HTML□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□Deep Web□□□□□□□Deep Web□□□□“□□□□□”(Ivisible Web)□□“□□”(Hdden Web)□□□□□□□□□□□□——Web□□□□□□□□□□□□□□□□□□□□

□□□□□□Deep Web□□□□□□□□□□□□WWW□□□□(Brgman 2001□He 2007□Raghavan 2001□□□□□□□□□(Mdhavan 2007□□□□□□□□□□□□□□□□deep-web□□□HTML□□□□□Deep Web□□□□□□□

[illegible]

[illegible]Deep Web[illegible]HTML[illegible]Web[illegible]deep-web[illegible]Deep Web[illegible]

[illegible]deep-web[illegible]Web[illegible]deep-web[illegible]Deep Web[illegible]

[illegible]Deep Web[illegible]Deep Web[illegible]HTML[illegible]deep-web[illegible](hst)[illegible]

[illegible](srfacing)[illegible]HTML[illegible]URL[illegible]URL[illegible]

[illegible]HTML[illegible]Web[illegible]Web[illegible]Deep Web[illegible]Web[illegible]get[illegible]HTML[illegible]post[illegible]

[illegible]deep-web[illegible]G[illegible]G[illegible]1000[illegible]Deep Web[illegible](Mdhavan 2008[illegible](Mdhavan 2009[illegible]

[illegible]deep-web[illegible]

[illegible]Deep Web[illegible]

[illegible]deep-web[illegible](mdiator)[illegible](mster)[illegible]Web[illegible]

[illegible]Web[illegible]Web[illegible]100[illegible]

[illegible]HTML[illegible]URL[illegible]URL[illegible]HTML[illegible]deep-web[illegible]Web[illegible]deep-web[illegible](sippet)[illegible]deep-web[illegible]

□□□□□□[1]□□□□□□□□□□□□HTML□□URL□Web□□□□□□□□□□□URL□□□□□□URL□□□Web□□□□□□□□□□□□□□□HTML□□□□□Web□□Cars.com□□□□□□□□□□5□□□□□□□□□□□□□□□□□URL□□2.4□□□□□□□□□□□□65□□□□(htp□//www.cars.com)□□□□□□□□□□URL□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

□□□HTML□□9-1□□□□□□□□□□□□□□□□□

□□“□□□□□”(iformativeness test)□□HTML□□□URL□□(Brbosa 2004□Ntoulas 2005□□

HTML□□□□□□□

□□□□□□□□□□□□□□□□□□□□□□□□□□HTML□□□□□□□(htp□//www.w3.org/TR/html401/interact/forms.html)□

HTML□□□form□□□□□□□9-1□□□□action□□input□□□□□□□□□□□□□□□□□□□□□□□□□□□□□(txt box)□□□□□□(slect menu□□□□□□select□□□□□□□□□□(ceckbox)□□□□□(rdio button)□□□□□(sbmit button)□□□□□□□□□□name□□□□□□□□□□□□HTML□□□(hdden input)□□

□9-1□□□□□□□□□□□□□HTML□□

```
□form action="http□//jobs.com/find"method="get"□
□input type="hidden"name="src"value="hp"/□
Keywords□□input type="text"name="kw"/□
State□□select name="st"□□option value="Any"/□□option value="AK"/□
□option value="AL"/□……□/select□
Sort By□□select name="sort"□□option value="salary"/□
□option value="startdate"/□……□/select□
□input type="submit"name="s"value="go"/□
□/form□
```

□□□□□□□□□□□□□□□get□□post□□□□□□□□□□□□HTTP□□□□□□□□□□get□□□□□□□□□action□□□□□□HTTP□□□□□□URL□□□□□□□9-1□http□//jobs.com/fnd?src=hp□kw=chef□st=Any□sort=salary□s=go)□□□□□post□□□□□□□HTTP□□□(bdy)□□□□URL□□□□action□□□9-1□http□//jobs.com/find)□□□□□get□□□□□□□□URL□□□□□□□□□□□□□□□□□□post□□□□URL□□□□□

[illegible]Web[illegible]URL[illegible]post[illegible]URL[illegible]

[illegible]HTML[illegible](sde effect)[illegible]post[illegible]

[illegible]get[illegible]Web[illegible]

[illegible]

[illegible]Javascript[illegible]onselect[illegible]onclick[illegible]onsubmit[illegible]Javascript[illegible]Javascript[illegible]Javascript[illegible]SpiderMonkey[illegible]V8[illegible]“[illegible]”[illegible]Web[illegible]deep-web[illegible]Javascript[illegible]deep-web[illegible]Web[illegible]get[illegible]“[illegible]URL”(iempotent URL)[illegible]

[illegible]

[illegible](rstricted language)[illegible]

[illegible]

[illegible]9-1[illegible]kw[illegible](kyword)[illegible]st[illegible]

(sate)[illegible]

[illegible]

[illegible]Any[illegible]

[illegible]

[illegible]HTML

[illegible]9-1[illegible]

[illegible]SQL[illegible]D[illegible]

[illegible]m[illegible]select*from D where P[illegible]

P[illegible]

[illegible]9-1[illegible]job[illegible]Jobs(position[illegible]city[illegible]state[illegible]desc)[illegible]

[illegible]select*from Jobs

where state='CA'and desc like'%chef%'[illegible]9-1[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

(bnding input)[illegible](fee input)[illegible]

[illegible]SQL[illegible]select*from D

where P'[illegible]P'[illegible]9-1[illegible]

[illegible]9-1[illegible]

表9-1：查询模板例子，每个模板都有不同的绑定输入项以及相应的表单提交的URL集合和底层Jobs数据库上的SQL查询

绑定输入项	表单提交URL和SQL查询
st	URL: *http://jobs.com/find?src=hp&kw=&st=S&sort=salary&s=go* 查询: select * from Jobs where state = **S**
kw	URL: *http://jobs.com/find?src=hp&kw=K&st=Any&sort=salary&s=go* 查询: select * from Jobs where desc like '%**K**%'
st, kw	URL: *http://jobs.com/find?src=hp&kw=K&st=S&sort=salary&s=go* 查询: select * from Jobs where state = **S** and desc like '%**K**%'

[illegible]SQL[illegible]

[illegible]deep-web[illegible]

·[illegible]

·[illegible]

[illegible]cityName)[illegible]state)[illegible]salaryMax[illegible]salaryMin)[illegible]

[illegible]

[illegible]

[illegible]10[illegible]

[illegible]Web[illegible]URL[illegible]

[illegible]deep-web[illegible]

[illegible]Web[illegible]Honda[illegible]Honda[illegible]URL[illegible]

[illegible]Deep Web[illegible]Web[illegible]URL[illegible]

[illegible]Web[illegible]Web[illegible]Web[illegible]

[illegible]

[illegible]

[illegible]URL[illegible]

[illegible]URL[illegible]URL[illegible]

[illegible]URL[illegible]

[illegible]1[illegible]2[illegible]URL[illegible]

[illegible]9-1[illegible]T1[illegible]st(State)[illegible]T2[illegible]sort(Sort by)[illegible]9-2[illegible]

表9-2：通过分析模板的表单提交生成的内容的相似性来确定模板是否为富信息量（informative）

查询模板 （绑定输入项）	模板表单提交 （富信息量的/非富信息量的）
T_1(st)	*http://jobs.com/find?src=hp&kw=&st=Any&sort=salary&s=go* *http://jobs.com/find?src=hp&kw=&st=AK&sort=salary&s=go* *http://jobs.com/find?src=hp&kw=&st=AL&sort=salary&s=go* … 网页内容不同→模板是富信息量的
T_2(sort)	*http://jobs.com/find?src=hp&kw=&st=Any&sort=salary&s=go* *http://jobs.com/find?src=hp&kw=&st=Any&sort=startdate&s=go* … 网页内容相同→模板是非富信息量的

□□□□□□□T1□□□□□□□□□□□□□□job□□□□□□□□□□□□□□□□□□Web□□□□□□□□□□□□□□□□□□□Web□□□□□□□□□□□□□□□□□□□□□□□□□□□□□T2□□□□□□□□□□□□□□□□□□□job)□□□

□□□□□□□□□□□□□□□□□Web□□□□□□□□(dstinctness)□□□□□□□□□□□□□□□□□□Web□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□Web□□□□□□

□□□□□□□□□□□□□□□□Web□□□□□□□□□□□□□□□□□1□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□2□□□3□□

□□□□□□

□□□□□□□□□□□□□URL□□□□□□□Web□□□□□□□□□□□□□□□□□□Web□□□□□□□□□□□□□□□□Web□□

□□□□□□□□□τ□□□□□□□□□□□□□□□□□□□Sd□□□□□□□□□St□□□□□□□□□□□□□|Sd|/|St|□τ□□□□T□□□□□□□□

□□HTML□□□□□□□□□□□□□□□□□□□□□Web□□

□□

□□

□94107□□□□□□□□Honda Civic□□□□□□□□□□□□□□□□□“□94107□□

□□□□Honda Civic□□□□□”(N search results for Red Honda Civic

in 94107□□□□□□□□□□{color make model zip}□□□□□□□□□□□□□

□□“No search results for{color make model}in{zip}”□□□□□□

□□

□□□□□□□

□□

□□□□□□□

□□□□□□□□□□

□□

□□□□□□□□□□□□□□□□□□□□□□□□□□□□n□□□□□□□2^n-1□□□□□□□□□□□

□□

□□□□□□□□□□□□□□□□

□□

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□T□□□k□□□□□□

□□□□□□□k□□□□□k-1□□□□□□□□□□□□□□T□□□□□□□□□□□

□□□□□□□□□1□□□□□□□□□□□□□□□□□□□□□□□□□1□□□□□□□□□□□□

□□□□□□□□□□□□1□□□□□□□□□□□□□□□□□□□□□2□□□□

[illegible]2[illegible]3[illegible]

[illegible]1[illegible]make[illegible]color[illegible]make[illegible]color[illegible]color[illegible]make[illegible]color[illegible]

[illegible]URL[illegible]URL[illegible]

[illegible]URL[illegible]URL[illegible]——[illegible]URL[illegible]

[illegible]

[illegible]HTML[illegible]

[illegible]

[illegible](tped input)[illegible]

[illegible]

[illegible]

[illegible]

[illegible](ste-specific)[illegible]

[illegible]Web[illegible]

[illegible](Brbosa 2004[illegible]Callan 2001[illegible]Ipeirotis 2002[illegible]Ntoulas 2005[illegible]

[illegible](ste-aware)[illegible]

[illegible]

[illegible]·[illegible]

·[illegible]Web[illegible]TF-IDF[illegible](Slton 1983[illegible]

[illegible]·[illegible]HTML[illegible]

[illegible]·[illegible]

[illegible]back-off[illegible]URL[illegible]

[illegible]Web[illegible]URL[illegible]deep-web[illegible]

[illegible]

[Doan 2001]

[1]

[illegible]

[illegible]Deep Web[illegible]Web)[illegible]

[illegible]Deep Web[illegible]MinPrice[illegible]MaxPrice)[illegible]

□□□□

[1]Barbosa,L. and J.Freire."Siphoning Hidden-Web Data through Keyword-Based Interfaces."SBBD 2004□309-321.

[2]Bergman,M. K."The Deep Web:Surfacing Hidden Value."Journal of Electronic Publishing□2001.

[3]Callan,J. P.and M.E.Connell."Query-based sampling of text databases."ACM Transactions on Information Systems□19□2□□97-130□2001.

[4]Doan,A.□P. Domingos,and A.Y.Halevy."Reconciling Schemas of Disparate Data Sources:A Machine-Learning Approach."SIGMOD Conference 2001□509-520.

[5]"Forms in HTML documents."http□//www.w3.org/TR/html401/interact/forms.html.

[6]He,B.□M. Patel,Z.Zhang,and K.C.-C.Chang."Accessing the Deep Web:A survey."Communications of the ACM□50□5□□95-101□2007.

[7]Ipeirotis,P. G.and L.Gravano."Distributed Search over the Hidden Web:Hierarchical Database Sampling and Selection."VLDB 2002□394-405.

[8]Madhavan,J.，L. Afanasiev,L.Antova,and A.Y.Halevy."Harnessing the Deep Web:Present and Future."CIDR 2009.

[9]Madhavan,J.，S. Jeffery,S.Cohen,X.Dong,D.KO'C.Yu,and A.Y.Halevy."Web-scale Data Integration:You can only afford to Pay As You Go."CIDR 2007.

[10]Madhavan,J.，D. KO'L.Kot,V.Ganapathy,A.Rasmussen,and A.Y.Halevy."Google's Deep-Web Crawl."PVLDB 1（2）：1241-1252（2008）.

[11]Ntoulas,A.，P. Zerfos,and J.Cho."Downloading textual hidden web content through keyword queries."JCDL 2005，100-109.

[12]Raghavan,S. and H.Garcia-Molina."Crawling the Hidden Web."VLDB 2001，129-138.

[13]Salton,G. and M.J.McGill.Introduction to Modern Information Retrieval.New York:McGraw-Hill，1983.

[14]SpiderMonkey(JavaScript-C）Engine,http：//www. mozilla.org/js/spidermonkey/.

[15]V8 JavaScript Engine,http：//code. google.com/p/v8/.

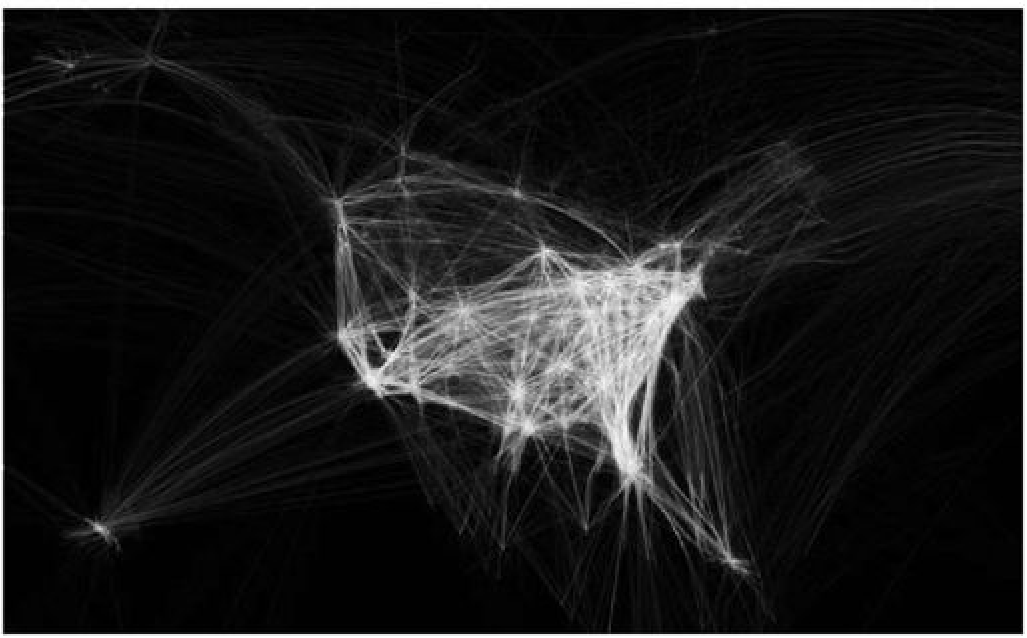
□10□ □□Radiohead□“House of Cards”

Aaron Koblin□Valdean Klump

□□□□□□□□□□(Gammy)□□□□□□Radiohead[1]□□□□□“House of Cards”□□□Gcode□□□□(htp□//code.google.com/radiohead)□□□3D□□□□□□□Flash□□□□□□□□□□□□□□□□□□□□□□□“Making of”□□□□□□□□□□□□□□□□□□□

□□□□□□□□□

2007□9□□□□□□□□□James Frost□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□James□□□□□□□□□□□□Coldplay□Norah Jones□Pearl Jam□□□□□□□□□□□□□□□□□□□□□□□□□“□□□□”(Fight Patterns)□□□□□10-1□——□□□□□□□GPS□□□□□□□□□□□□□□□□□——□□□□□□□□□□□□□□□□□□□□□



□ 10-1□□□“□□□□”□□□□□□2005□□□□□22□

□□□□□James□□□□□□Justin Glorieux□□□Proccessing□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□Interpol□□□□□“Rest My Chemistry”□□□□□□□2008□3□□□□□□□□□□□□□□□Processing□□□□□□□□□□□□□□□□http□//processing.org/□□

[illegible]Processing[illegible]

[illegible]

[illegible](ULA)[illegible](CNS)[illegible]CENS[illegible]James[illegible]“[illegible]”[illegible]Radiohead[illegible]

[illegible]GCode[illegible]

[1]Radiohead[illegible]“[illegible]”[illegible]Thom Yorke[illegible]Ed O'Brien[illegible]Jonny Greenwood[illegible]Colin Greenwood[illegible]Phil Selway[illegible]20[illegible]90[illegible]/[illegible]

□□□□□□

□□□□□□□□□□□□□□□□“House of Cards”□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□Thom Yorke□□□□□□(pint clouds)□□□□□□□□□Thom□□□□□□□□□□□□□□□□□□□□□□□□Velodyne□□□Lidar□□□□□□Geometric Informatics□□□□□□□□

Velodyne□□□Lidar

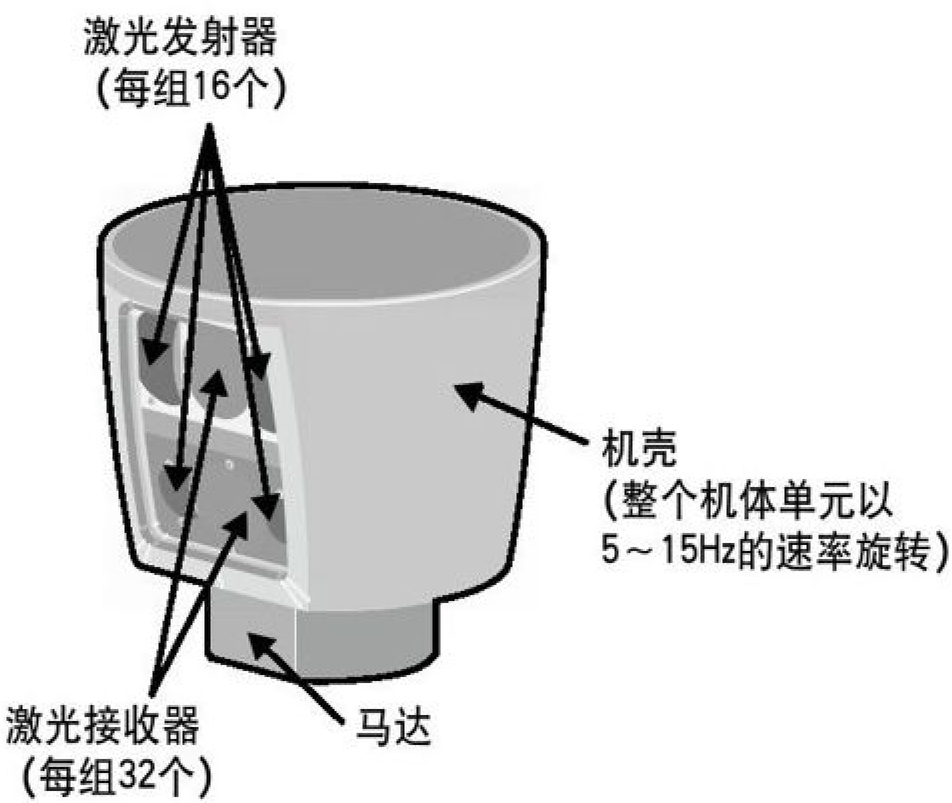
Velodyne□□□□□□□□San Jose□□□□□□□□□□□□□□□□□□□□□□□□□□□□Battle Bots□Robot Wars□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□“House of Cards”□□□□□□□□□□□□□□HDL-64E Lidar□□□□□□□□HDL-64E□□□□□□□□□□□□□□□□2007□DARPA□□□□□(Uban Challenge)□□

Velodyne□□□HDL-64E Lidar□□□□□□□□64□□□□□□64□□□□□□□□□□□□□□□□360°□□□□□26.8°□□□□□□□□□□□□□□□□□100□□□□□□□□□□5MB□□□□□□□□□□□Lidar□600RPM□10Hz)□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□300□900RPM□□□□□□□□□□□□□□□□□□□□□□□□□□□□900RPM□

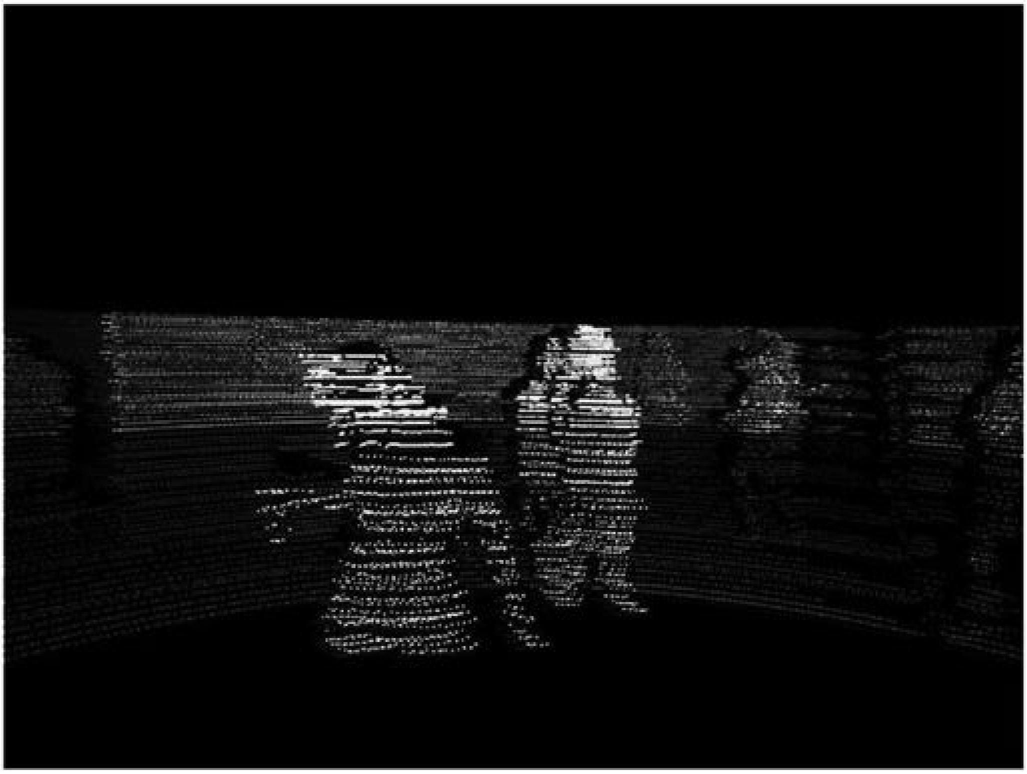
Lidar□□□□□□□□□□□□□□□□□□□□□□□□50□□□□□□□□□□□□□□□□□□□□□□□□120□□Lidar□□□□□□□3□□□□□□□□□□□□□□□□□□□□□□□□□□

[illegible]-[illegible]32[illegible](Lser bank)[illegible]10-2[illegible]Lidar[illegible]10-3[illegible]

Lidar[illegible]64[illegible]5ns[illegible]Lidar[illegible]



□ 10-2□Velodyne□□□Lidar□□□□□□Velodyne□□□



□ 10-3□□□□□□□□□□□□□□□Velodyne□Lidar□□□□□□□□□□□□□□
□□□□□□□□□□□□□□□□□□□□□□□□□23□

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□Avalanche Photodiode(APD)□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□APD□□HDL-64E□□□□□□□□□100□□□□□□□□□□□5MB□□□□□□

□□□□□□□□□□□□□□□□□100BaseT□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□600RPM=10Hz□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

[illegible]-[illegible]

[illegible]

Geometric Informatics[illegible]

Lidar[illegible]Thom Yorke[illegible]

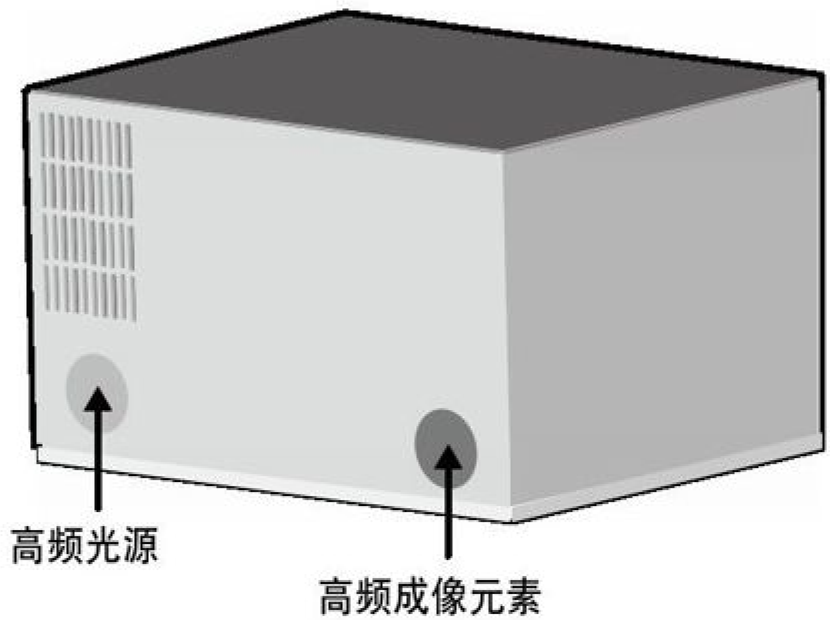
[illegible]2005[illegible]SIGGRAPH[illegible]Geometric Informatics[illegible]Geometric Informatics[illegible]GeoVideo[illegible]

GeoVideo[illegible]Lidar[illegible]0.2mm[illegible]Lidar[illegible]2cm[illegible]GeoVideo[illegible]Thom Yorke[illegible](pint cloud data)[illegible]GeoVideo[illegible]

[illegible]10-4[illegible]60[illegible]54MB[illegible]180[illegible]

[illegible]GeoVideo[illegible]“[illegible]”[illegible]

[illegible]



[illegible] 10-4[illegible]Geometric Informatics[illegible]Geometric Informatics[illegible]24[illegible]

GeoVideo[illegible]3D[illegible]“House of Cards”[illegible]Thom Yorke[illegible]“[illegible]”[illegible]Thom Yorke[illegible]——[illegible]——[illegible]——[illegible]

[illegible]

[illegible] Lidar [illegible] GeoVideo [illegible]

[illegible]

[illegible] Lidar [illegible] “House of Cards” [illegible] “Lidar [illegible]” [illegible] GeoVideo [illegible] Lidar [illegible]

□□

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□Thom Yorke□□□□□□□□□□□□□□□□□□□□□□2067.csv□□□□□□□□□1067□□□□□□□□30□□□□□□□□□1001.csv□□□□□□□□□□□□□□36□□□□□□□□□GCode□□□□□□□□□□□□□□□□□□□□□

70.05□162.48□-79.32□122

70.23□165.26□-78.82□112

70.46□168.00□-77.55□95

□□□□□□x,y,z□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

□x,y,z□□□□□□□□GeoVideo□□□Lidar□□□□□□□□0□0□0□□□□□□□□□□□□□□□□□□70.46□□□□□x□□□0□□70.46□□□□□□□□□□□□□□□□□□□□□□□□0□0%□□□□256□100%□□□□

□GCode□□□□□□□□Thom□□□2000□□□□□□□□□□□□□□□□1□□□HoC_DataApplications_v1.0.zip□□

□□□□□□“□□”

2008□5□□□□□□□□□□□Palm Beach□□□□□□□□□□□□□□“House of Cards”□□□□□□□Radiohead□□GCode□□□□□□□□□□□□□“Making Of”□□□

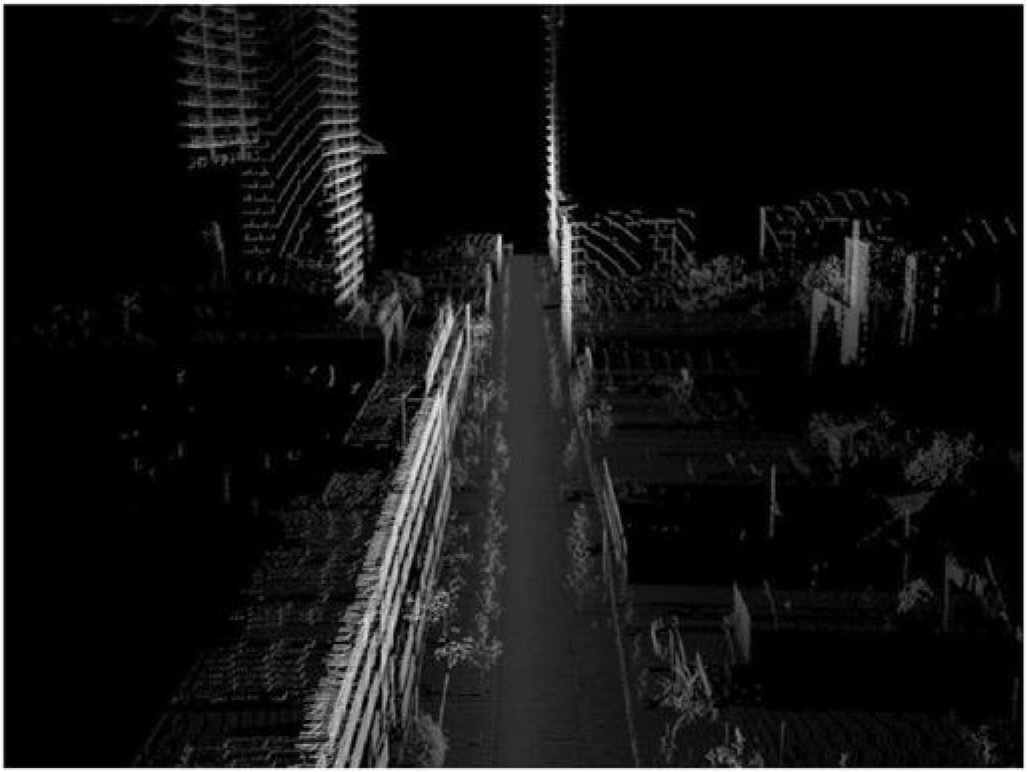
□□Lidar□□

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□Lidar□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

□□DARPA□□□□□□□□□□□□□□Lidar□□□□□□□□□□□□□□□90°□□900□□

□□□64□□

□10-5□□□Lidar□□□Lidar□□□□□□□□□□□



□　10-5□Velodyne Lidar□□□□□□□□□□□□□□□□25□

□□
□□□□□□□□□□□□□□□□□□□10mph□□□□□□□□□□□□□□□□□□□□□□□□□

□□□□□□□Lidar□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□
□□□Larry
Zience□□□Velodyne□□□□□□Rick Yoder□“□□□□”□□□□Rick
Yoder□□“□□□□
□”□□□□Rick□□□□□□□□□□□□□□□□Lidar□□□□□□□□□□□□□□□Larry
□“□”□□□□□□□□□□

□□□Rick□□□□□□□□□□□□□□□□□□□□□□

[illegible]“[illegible]20[illegible]45°[illegible]”[illegible]“[illegible]”

[illegible]10-6[illegible]“[illegible]”[illegible]“[illegible]”[illegible]



[illegible] 10-6[illegible]Velodyne Lidar[illegible]26[illegible]Lidar[illegible]

□□□□Lidar□□□□□□□□□□□□□□□□□□□□3□30□3□55□□□□□□□□□□

□□□□□□□□□□□□□□□□□□□□□64□Lidar□□□□□□□□□□□□□□□□□□□□□□

□□□□□——Lidar□□□□□□□□□□□□□□——□□□□□□□□□□□□□□□□□□□□□

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□Lidar□□□□□□□□□

□□□□□□□□

□□

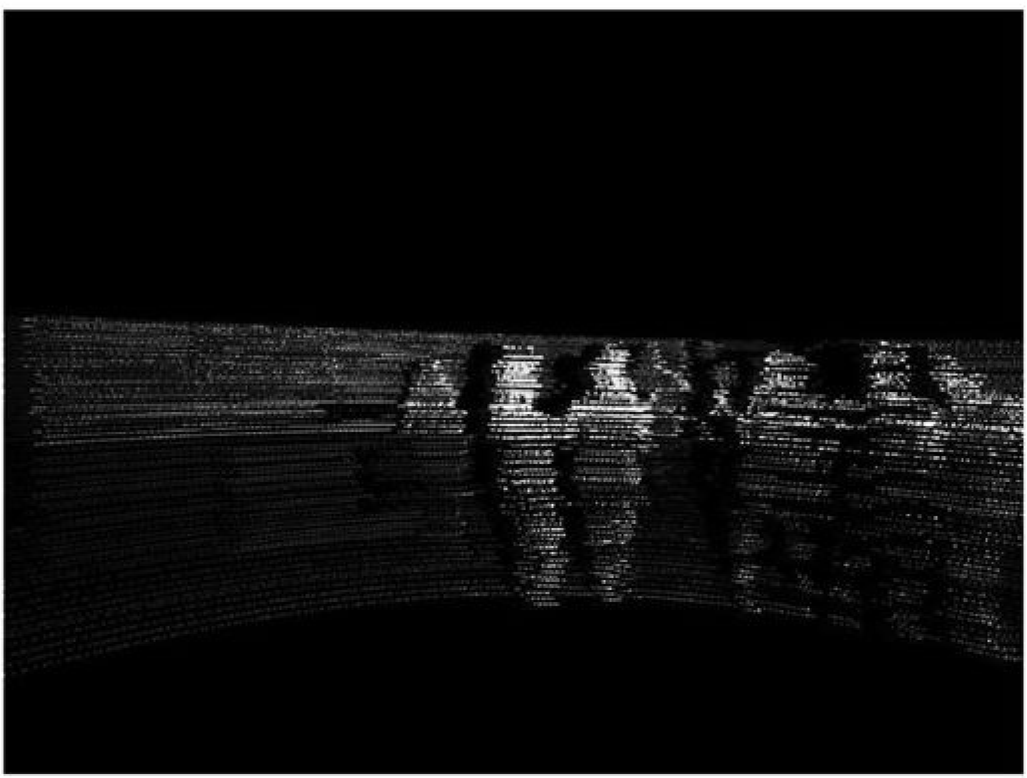
□□Radiohead□□

□□□□□□□□□□□□□

□□□□□□□10-7□□□□□□□□□□□□□□64□□□□□□□□□□□□□□□□□□□□□

Lidar□□□32□□□□□□□□□□□□□□□□□□□□□□Lidar□□□□□□□□□□□□□□□

□□□□□□□□□□□□□□□□□□□□□□□□□□



□ 10-7□□□□□□□□□□Velodyne Lidar□□□□□64□□□□□

Lidar□□□□□□□□□27□

□□□□Lidar□□□□□□□□□□□2cm□□□“□□”□□□Lidar□□□□□□□□Thom Yorke□□□□□□□□

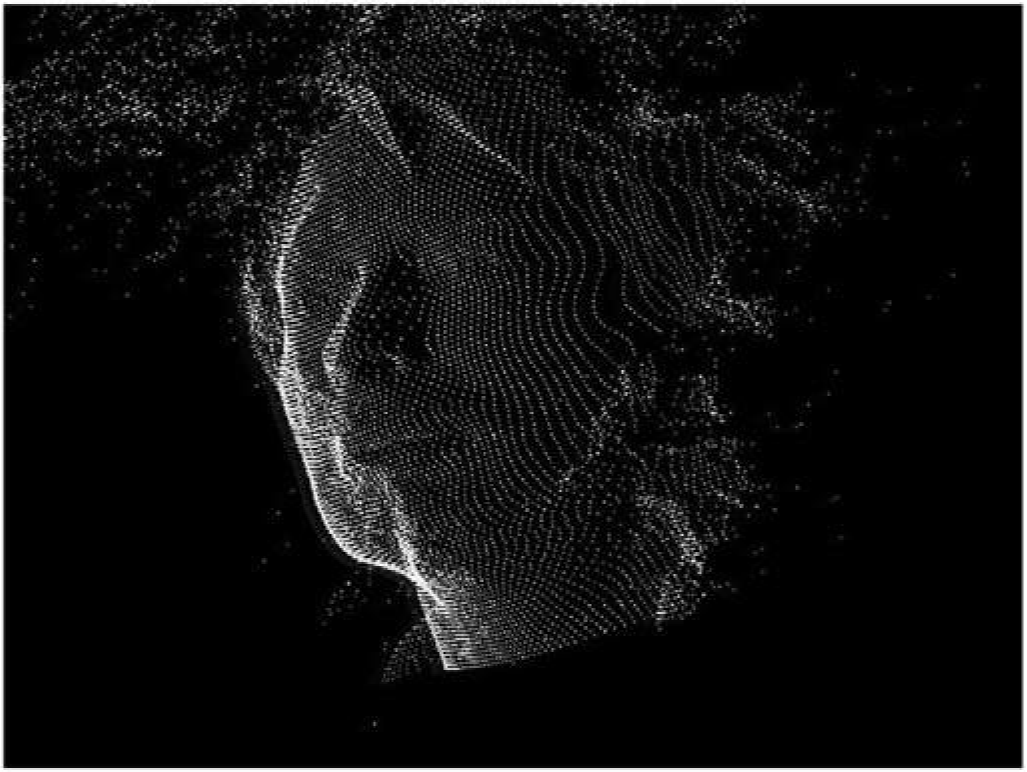
GeoVideo□□

□□GeoVideo□□

Thom Yorke□□□□□“□□”□□Lauren Maher□□□□□1□05□□□□□□□□□□□□□□□□Geometric Informatics□□GeoVideo□□□□

GeoVideo□□□□□□□□□□□□□□□□Geometric Informatics□□

□□□□□□□□□□□□□□□□□James Frost□□□□□□□Thom Yorke□□□□□□□□□□Thom Yorke□□□□□□□□□□10-8□□□□□□□□□□□□□□□Thom Yorker□□□□□□□□□□□□□□□□□□□□□



[illegible] 10-8[illegible]Thom Yorke[illegible]Radiohead[illegible]28[illegible]

[illegible]

[illegible]Thom[illegible]“[illegible]”[illegible]

[illegible]10[illegible]GeoVideo[illegible]Lidar[illegible]Thom[illegible]

[illegible]

[illegible]Lidar[illegible]510 Systems[illegible]Lidar[illegible]Pierre-Yves Droz[illegible]Lidar[illegible]

[illegible]DVD[illegible]Pierre[illegible]

[illegible]63[illegible]Lidar[illegible]x,y,z[illegible]

[illegible]Pierre[illegible]Lidar[illegible]Velodyne[illegible]Lidar[illegible]Pierre[illegible]Lidar[illegible]

[illegible]510 Systems[illegible]4GB[illegible]50GB[illegible].obj[illegible]

[illegible]

[illegible]The Syndicate[illegible]Santa Monica[illegible]1[illegible]

The Syndicate[illegible]Brandon Davis[illegible]

[illegible]Radiohead[illegible]“[illegible]”[illegible]“[illegible]”[illegible]

Brandon Davis[illegible]“[illegible]”[illegible]

[illegible]——[illegible]Lidar[illegible]Thom Yorke[illegible]

[illegible]Thom[illegible]Brandon[illegible]

[illegible]

[illegible]

[illegible]The Syndicate[illegible]2D[illegible]
3D[illegible]

[illegible]James[illegible]

“House of Cards”[illegible]

[illegible]Nicholas Wayman Harris[illegible]

□□□□

“House of Cards”□□□□G□□□□□□□□□□□□□□□□□□□□□2008□7□11□□□□G□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□3D□□□□□□□□G□□□□□□□□□□□□□□□□□□□

□□□□□□□□□□□□Aaron Meyers□□□Flash□□□Flash□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□Thom Yorke□□

□□□□□□□□□□□□□□□□□□□□□Processing□□□

□□□□□□□□□□□□□□□□□□□□□Processing□□□□□□□□□□□□□□□□□□□□□Thom Yorke□□□□□□□□

```
import processing.opengl.*□
int frameCounter=1□   //Declare a variable to store which frame we're dealing with
void setup□□{   //Here we set up the program
size□1024□768□OPENGL)□   //This is the render size.We'll use OpenGL to draw as
//fast as possible
//frameRate□30□□   //Uncomment to watch the animation at 30 frames per second.
strokeWeight□1□□   //Draw lines at a width of 1□for now.
}
void draw□□{   //Here we state the things we're going to do every frame
background□0□□   //We'll use a black background
```

```
    translate(width/2□height/2□□   //The data has 0□0□0 at the center and we
want to
    //draw that point at the center of our screen
    translate□-150□-150□□   //Let's adjust our center slightly
    scale□2□□//Let's draw things bigger
    //rotateY(frameCounter/50.0f□□   //If uncommented,this makes the data
rotate over
    //time
    //rotateY(mouseX/150.0□□   //If uncommented,this uses the mouse's
horizontal
    //location to adjust the rotation
    String[]raw=loadStrings(frameCounter+".csv"□□//Here we load the current
frame
    //data into an array
    for(int i=0□i□raw.length□i++□{   //Now we loop through each line of the
    //raw data
    String[]thisLine=split(raw[i]□'□'□□   //For each line we're going separate
    //each parameter
    float x=float(thisLine[0]□□   //Now we make a decimal variable for each
    //parameter
    float y=float(thisLine[1]□□
    float z=float(thisLine[2]□□
    int intensity=int(thisLine[3]□□
    stroke(intensity*1.1□intensity*1.6□200□255□□//We set the color of each point
to
    //correspond to the data's
    //intensity value
    line(x,y,z,x+1□y+1□z+1□□   //Here we draw a little line for each point□this
    //is much faster than a more complex object and//we'll be drawing a lot of
them
    }
    frameCounter++□//Add one to the frame variable to keep track of what frame
we're
    //currently on
    if(frameCounter□2101□{   //If we get to the end of the data we'll exit the
    //program
    exit□□□
    println□"done"□□
    }
    //saveFrame□"renderedFrames/"+frameCounter+".tga"□□   //This would be a
way to save out
    //a frame
    //*remember you're saving files to your harddrive□*
    }
```

[illegible]

[illegible]Thom Yorke[illegible]

```
rotateY(frameCounter/50.0f);
```

[illegible]Thom[illegible]y[illegible]

```
rotateY(mouseX/150.0);
```

[illegible]Thom[illegible]

[illegible]QuickTime Pro[illegible]Final Cut[illegible]After Effects[illegible]Youtube[illegible]“House of Cards”[illegible]

(htp[illegible]//www.youtube.com/group/houseofcards)[illegible]

[illegible]

[illegible]

[illegible]7[illegible]

1.[illegible]

[illegible]“House of Cards”[illegible]

2.[illegible]

[illegible]

3.[illegible]

[illegible]“House of Cards”[illegible]Velodyne Lidar[illegible]GeoVideo[illegible]Lidar[illegible]

4.[illegible]

[illegible]Thom[illegible]

[illegible]

5.[illegible]

□□□□GeoVideo□□□□□Thom Yorke□□□□□□□□□□□□□□□□□□□□□
□□□□□□□□□□□□□□□□□□□□□□□□□

6.□□□□□□

□□□□□□□□□□□□□□□□□□□□□□□“House of Cards”□□□□□□□□□□□
□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□“□□”□□□□□□□□□□
□□□
□□□10□□□GCode□□□□□□□□□□□□□□□□□□□

7.□Radiohead□□□□□□□

□□□
□□James Frost□
□□Velodyne□Geometric Informatics□510 Systems□The
Syndicate□□□□□□□□□□□□□□□□□□□□□□□□□□□□

□11□ □□□□□□□

Michal Migurski

[illegible]

[illegible]

[illegible]

[illegible]Crimespotting[illegible](htp[illegible]//oakland.crimespotting.org)[illegible]Stamen Design[illegible](htp[illegible]//stamen.com)[illegible]Crimespotting[illegible]11-1[illegible]——CrimeWatch(http[illegible]//gismaps.oaklandnet.com/crimewatch/[illegible]Crimespotting[illegible]Crimespotting[illegible]Stamen[illegible]Ben Cerveny[illegible]“[illegible]”[illegible]

[illegible] 11-1[illegible]Stamen Design[illegible]——[illegible]

Crimespotting[illegible]

[illegible]“[illegible]”[illegible]

□□

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□/□□□
□Adrian Holovaty□2005□□□□Chicago Crime□□
(htp□//chicagocrime.org)□Chicago Crime□□□GMap□□□□□□□
□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□GMap
API(http□//code.google.com/apis/maps/□□□□□□□□□□□□□□□
□□□
□Holovaty□□□□□□□□□□□□□□□□□□□□□“□□”□□□□□□□□□G□□
2005□□□□□□□□□□□□□□□□□□□□□□□□GMap□□□□□□□□□□□
□□□□□□□□□□□□□□□□□□□□□□□□□□□□

Chicago Crime□□□□□□□□□□□□□□□□□Paul Rademacher□
□□Housing Maps(http□//housingmaps.com)□□□□□□Chicago
Crime□□□□□□□□□□□□□□□□□□□□□□□□Michael Frumin□
Jonah Perretti□New York□□□□Eyebeam□□□FundRace□□
(htp□//fundrace.org)□□□2004□□□□□□□□□□□□□□Rich
Gibson□Schuyler Erle□Jo Walsh□□□□□Mapping Hacks□
□O'Reilly□□□□□□□□□□□□□□Web□□□□□□□□□□□□□□□□□□□□
□□□
□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□Steven Johnson□
□□“□□□□□”(Te Long Zoom)□□□□□□□□□□□□□□□□□

GMap[illegible]Web[illegible]GMap[illegible]

[illegible]Jesse James Garrett[illegible]——Ajax[illegible]JavaScript[illegible]XML)[illegible]Ajax[illegible]HTML[illegible]JavaScript[illegible]Web[illegible]

[illegible]Web[illegible]

[illegible]CrimeWatch[illegible]“[illegible]”[illegible]

□□□□□□

□□□□CrimeWatch□□CrimeWatch□□“□□□□□”(Wzard)□□□CrimeWatch□□□□□□□□□□

1.□□□□□□□□□□□□□□□□□□□

2.□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

3.□□□□□□□□□□□□□□□□□

CrimeWatch□□□□□□□□□□□□□□□□□□□□□□□□□□

□□□□□□□□□□□□□□□□□□□□□□□CrimeWatch□□(golocation)□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

□□“□□”(hok)□“□□”□□□□□□□□□□□□□□CrimeWatch□□□HTTP Cookies□□□(Tken)□□□□□□□□HTTP□□HTTP□□□□Charles(http□//charlesproxy.com/□□Mozilla□□LiveHTTPHeaders(http□//livehttpheaders.mozdev.org/□□□□□□

[illegible]HTML[illegible](pge-scraping)[illegible]

Leonard Richardson[illegible]BeautifulSoup[illegible]

(htp[illegible]//www.crummy.com/software/BeautifulSoup/[illegible]

BeautifulSoup[illegible]HTML“[illegible]”[illegible]

[illegible]HTML[illegible]JavaScript[illegible]

[illegible]/[illegible]

[illegible]Unix[illegible]Shell[illegible]

cURL(http[illegible]//curl.netmirror.org/[illegible]1[illegible]

[illegible]HTTP[illegible]URL[illegible]CGI[illegible]

Post[illegible]

CrimeWatch[illegible]Cookie[illegible]HTTP[illegible]Cookie[illegible]

Jar[illegible]CrimeWatch[illegible]JavaScript[illegible]

[illegible]HTML[illegible]

[illegible]Web[illegible]HTML[illegible]

[illegible]“[illegible]”User-Agent[illegible]IE[illegible]Mozilla[illegible]

User-Agent[illegible]

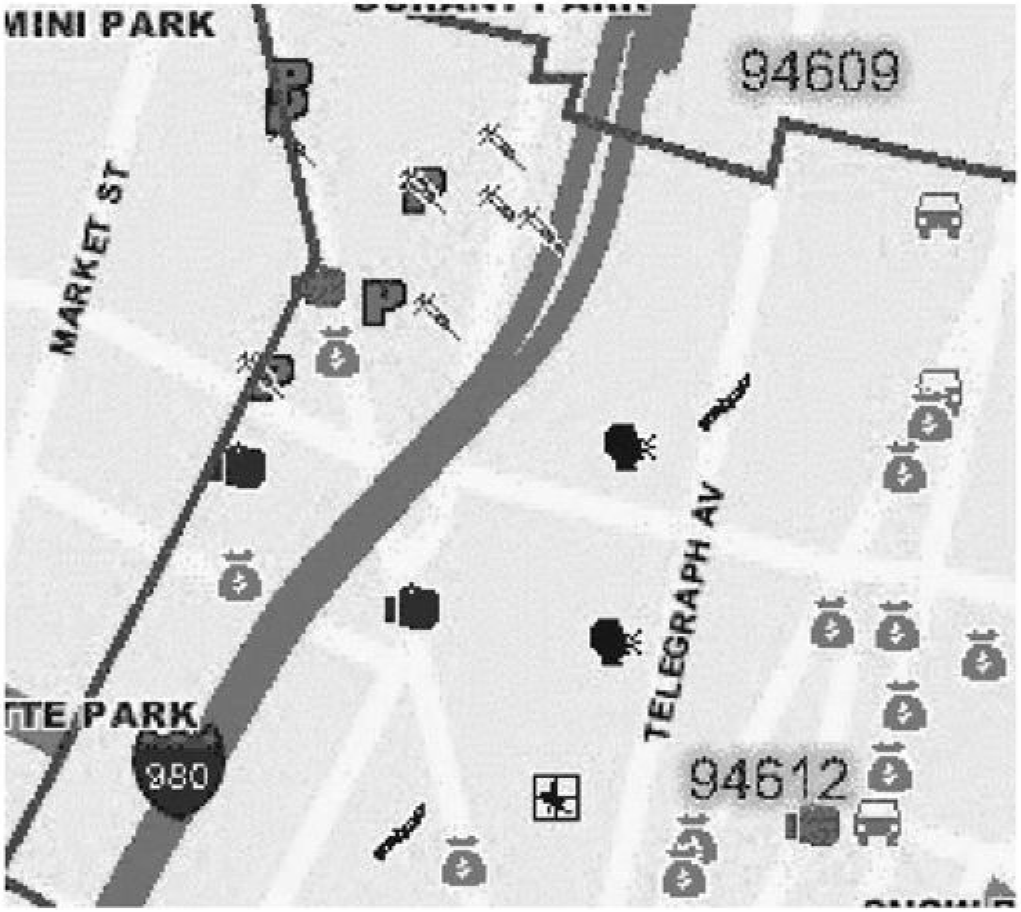
[illegible]

[illegible]“[illegible]”[illegible]CrimeWatch[illegible]

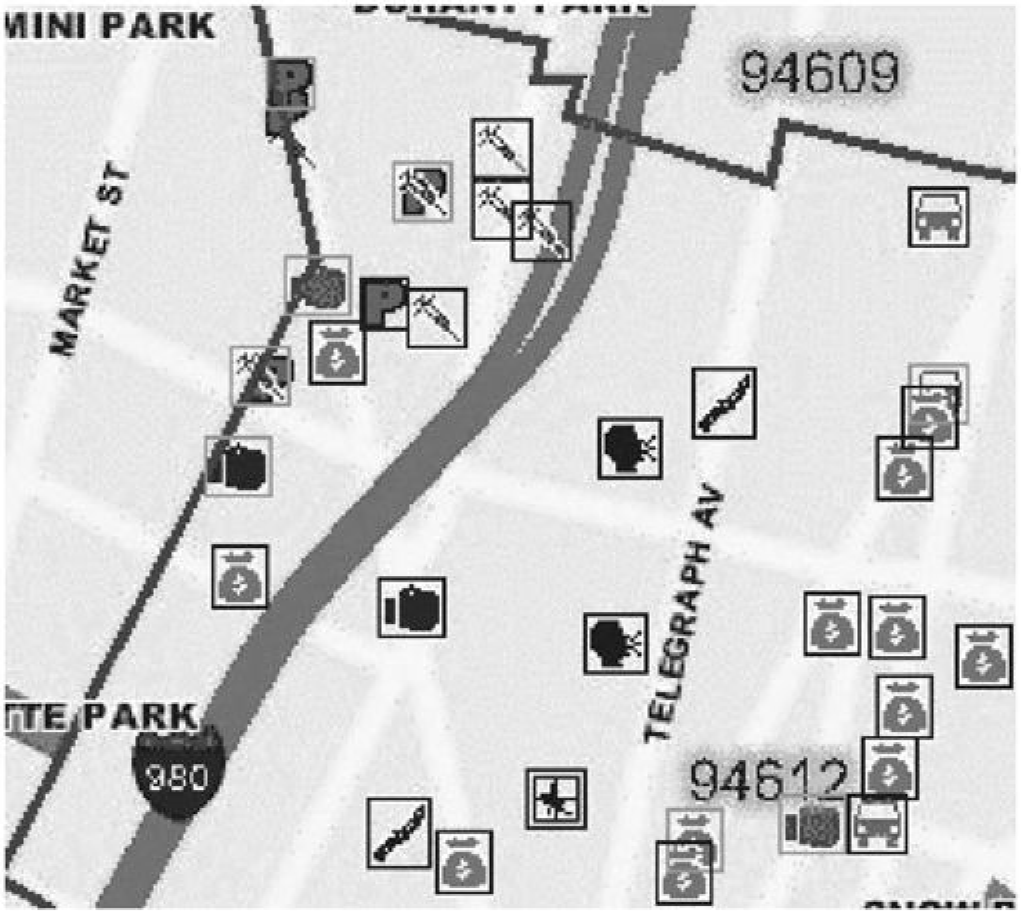
[illegible]

NumPy(http□//numpy.scipy.org/□□□□□□□□□□□□□□□□Python□□□

□□□□11-2□□11-3□□□CrimeWatch□□□□□□□□□□□□□□□□□□□□□□



□　11-2□CrimeWatch□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□29□



[illegible] 11-3 [illegible] CrimeWatch [illegible] 30 [illegible]

[illegible] CPU [illegible]

[illegible] 11-4 [illegible] 11-2 [illegible]



[illegible] 11-4[illegible]CrimeWatch[illegible]

[illegible]

[illegible]

[illegible]CrimeWatch[illegible]3[illegible]
(Plice Service Area)[illegible]
[illegible]CrimeWatch[illegible]
[illegible]

□□□□□□□CrimeWatch□□□□□□□□□□□□□x,y□□□□□□□□□□Simon Willison□GetLatLon□□(htp□//getlatlon.com)□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□6□□□□□□□□□□□□□□□□□□□□□18□□11-5□□



□ 11-5□□□□□□□□□□□□□□□□□□□□□□□□□□□□31□

□□100□□□

□□□□□□□□□□□□□□□□□□□□□Twisted Python(http□//www.twistedmatrix.com/□□□□□□□□□□□□□□□□□□□□□□□□□□□□CrimeWatch□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

[illegible]

[illegible]CrimeWatch[illegible]

[illegible]8[illegible]13[illegible]6[illegible]

[illegible](Pge Scraper)[illegible]CrimeWatch[illegible]

[illegible]/[illegible]——[illegible]

[illegible]——[illegible]CrimeWatch[illegible]

[illegible]6[illegible]

[illegible]CrimeWatch[illegible]“[illegible]”[illegible]

[illegible]

[illegible]

[illegible]2007[illegible]

[illegible]

[illegible]

2007[illegible]8[illegible]2[illegible]Oakland Post[illegible]Chauncey Bailey[illegible]

[illegible]Bailey[illegible]

[illegible]

[illegible]Oakland Tribune[illegible]Sean Connelly[illegible]Katy Newton[illegible]

[illegible]Not Just A Number[illegible]

(htp[illegible]//www.bayareanewsgroup.com/multimedia/iba/njn/[illegible]

Connelly[illegible]Newton[illegible]Not Just A

Number[illegible]

[illegible]

[illegible]Web[illegible]

Stamen[illegible]Tom Carden[illegible]Flash[illegible]

Eric Rodenbeck[illegible]

[illegible]Crimespotting[illegible]11-

6[illegible]

Lake Merritt[illegible]“[illegible]”(Sippy Map)[illegible]

[illegible]Berkeley/Emeryville[illegible]

Piedmont[illegible]San Leandro[illegible]City

of Oakland[illegible]

[illegible]Crimespotting[illegible]“[illegible]”[illegible]

[illegible]Peter Morville[illegible]

[illegible]——“[illegible]”(fndability)[illegible]

[illegible]Web[illegible]2005

[illegible]GMap[illegible]New York

Times[illegible]

[illegible]“[illegible]”[illegible]

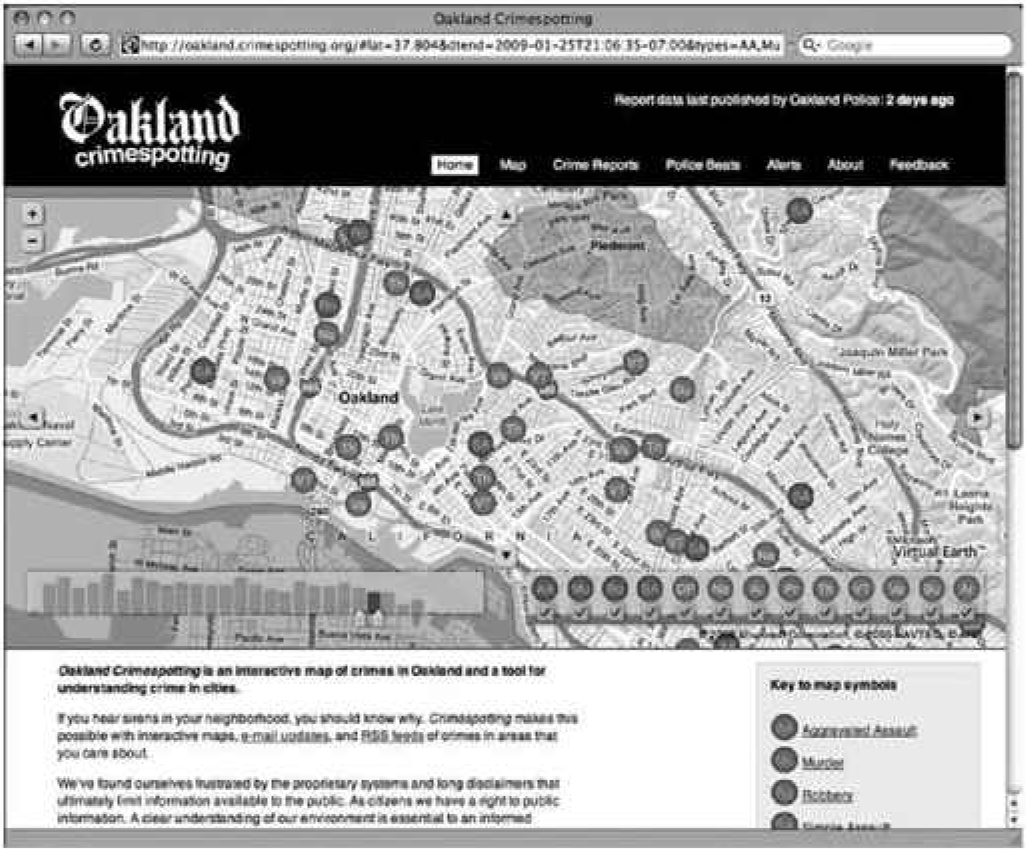
[illegible]“[illegible]”(Sented Widgets)[illegible]

[illegible]Wesley Willett[illegible]Jeffrey Heer[illegible]Maneesh

Agrawala[illegible]2007[illegible]

(htp[illegible]//vis.berkeley.edu/papers/scented_widgets/2007-

ScentedWidgets-InfoVis.pdf)[illegible]



□　11-6□□□□Crimespotting□□□□□□□□□□□□□□□□□□□□□□□□□□□32□

□□□□□□□“□□”□□□□□□□□□□□□□□□□□□□Web□□□□□□□□□□□□□□□□□□□□□□□“□□”□□□“□□□□”□□

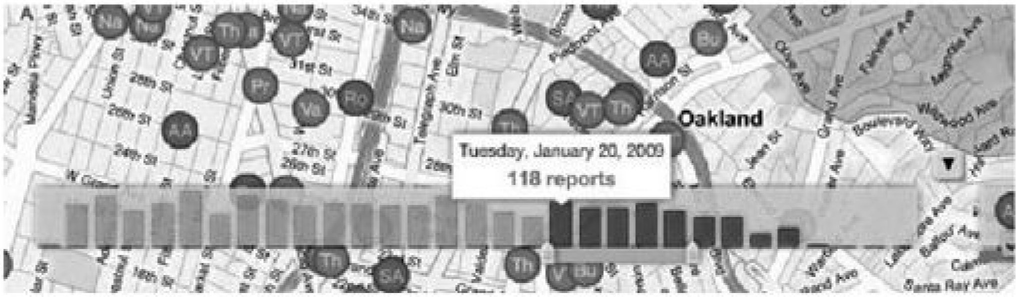
Crimespotting□□11-7□□□11-8□□Measure Map(http□//measuremap.com/□□□□□□□□□□□□

Measure Map[illegible]Adaptive Path[illegible]Jeffrey Veen[illegible]G[illegible]

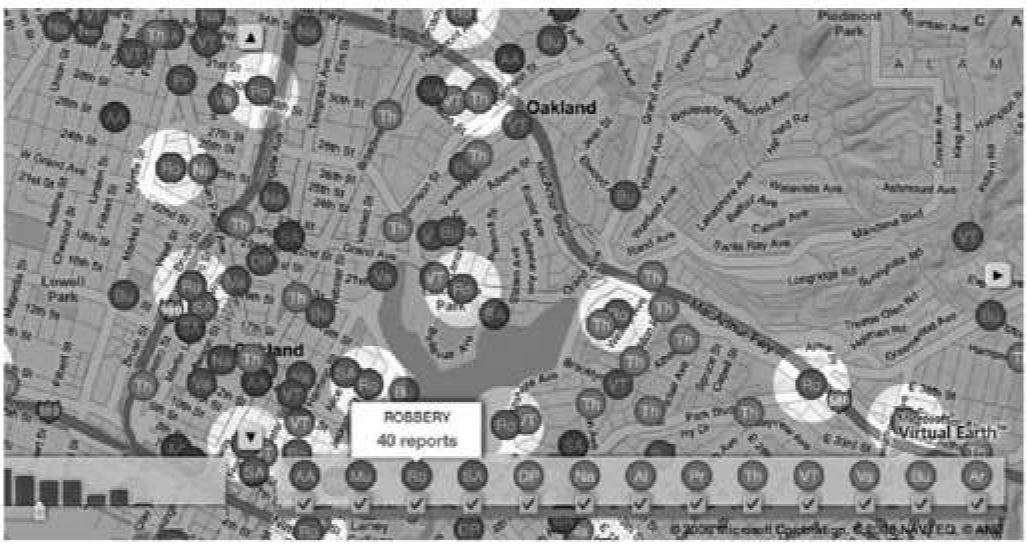
[illegible]Measure Map[illegible]Flickr[illegible]

[illegible]

[illegible]



[illegible] 11-7[illegible]Crimespotting[illegible]33[illegible]



[illegible] 11-8[illegible]34[illegible]

[illegible]11-8[illegible]

[illegible]Max OS X[illegible]

[illegible]

[illegible]

[illegible]

□□Flash□□JavaScript□□□OpenLayers(http□//www.openlayers.org)□

□□BSD□□□□□Modest Maps(http□//www.modestmaps.com/□□Modest Maps□□□Stamen□□□□□□□□□□□□□□□□

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□Crimespotting□□□□□□URL“□□”□

·□□□□http□//oakland.crimespotting.org□□□□□□http□//oakland.crimespotting.org/map□

·□□□□□□□http□//oakland.crimespotting.org/crimes,http□//oakland.crimespotting.org/crimes/Robbery,http□//oakland.crimespotting.org/crimes/2009-01-09□http□//oakland.crimespotting.org/crimes/2009-01-09/Robbery□

·□□□□□□□http□//oakland.crimespotting.org/crime/2009-01-09/Robbery/113569□

·□□□□□http□//oakland.crimespotting.org/beat/04X□

□□□URL□□□□□□□□□□□□□□□□□□□□URL□□□□□Matt Biddulph□2005□□□□□“Designing Data For Reuse”(htp□//www.hackdiary.com/slides/xtech2005/□□□“1□9□□□□”□“04X□□”□□□□□□□□□□——□□□□□□“/□□/2009-01-09/□□”□□□□“/□□/□□/2009-01-09”□□□□□□“/□□(cime)/□□”□□□□□□“/□□□□(cimes)/□□”——□□□□HTTP□□URL□□□□□□□□□□□□□□□□URL□□□□□□□□□□□□□□□□□□□□PostgreSQL□□□Josh Berkus□□□□□□□□□□□“Primary Keyvil”□□□□□□□□□□□□□(htp□//it.toolbox.com/blogs/database-soup/primary-keyvil-part-i-7327□□

□□□□□□□□□□□□□□□□“id”□□id□□

□□□08-056061□□□ (htp□//oakland.crimespotting.org/crime/2008-08-

01/murder/93014□□□□□□9□□2008□8□□“ASSAULT W/SEMIAUTOMATIC FIREARM ON PEACE OFFICER/FIREFIGHTER”□□□□□□□□□□□□□□□□□□□□□□□URL□□□□□□□□□□□□□□□ID□□□□

□URL□□□□□□□□□□□□□□□□□□□□□□□□□□□CrimeWatch□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□(Wzard)□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□10□□□□□□□□URL□□□□□□□□□□□□□□□□□□□□Leonard Richardson□□□□□□□□□□□□□URI□□□WWW□20□□□□□□□□□□□□□□□□□□□□□□Richardson 2008□□□□□□“Justice Will Take Us Millions Of Intricate Moves”(htp□//www.crummy.com/writing/speaking/2008-QCon/□□□□□□Web□□□□□□□□□□□□□□□URI□□□□□□□□HTTP□□□□□□□□□□□□□□□□□□□□□□□HTML□□□□□□□□□□□□□□□□□□□□URI□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□HTML□□□□□□□Web□□□□□□□HTML□□Roy T.Fielding□2000□□□□□□□□□□“□□□□□□(Rpresentational State Transfer,REST)□□□□□□”(htp□//www.ics.uci.edu/□fielding/pubs/dissertation/top.htm)□□□□□□□□□□□□□□□□□□□□□□□□□□□□Crimespotting□□□□Falsh□□□□□□□□□□□□1993□□Web□□□□□□□□□API□Flash□RSS□Atom□Feed□□□□XML□□□□□□□□□□□□□CSV□□□□□□□□□□□□□□□□□□□□□□□□□API□

2007[illegible]8[illegible]CrimeWatch[illegible]

[illegible]

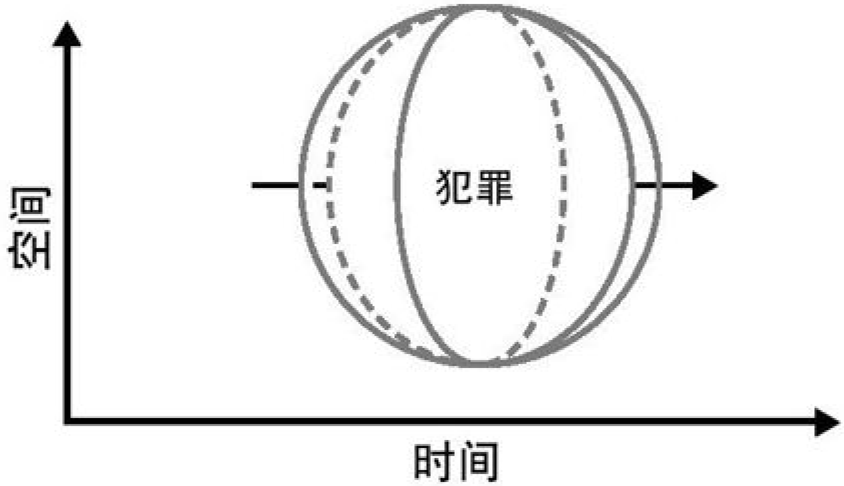
[illegible]

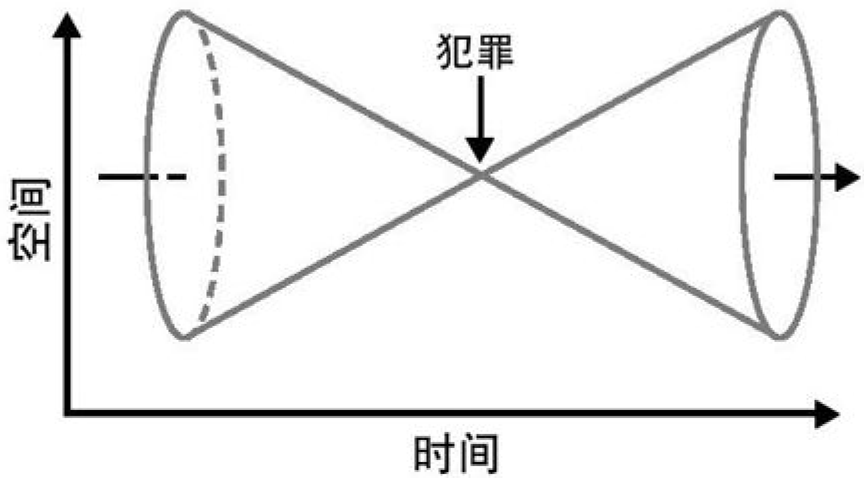
[illegible]

[illegible]CrimeWatch[illegible]

[illegible]Adam Greenfield[illegible]“[illegible]”[illegible]11-9[illegible]-[illegible]0.25[illegible]

Greenfild[illegible]“[illegible]”(lght cone)[illegible]-[illegible]





11-9

-

San Pablo Avenue

International Boulevard

[illegible]

[illegible](pge scraping)[illegible]JavaScript[illegible]CrimeWatch[illegible]IP[illegible]

[illegible](bat-specific)[illegible]Lake Merritt[illegible]http[illegible]//oakland.crimespotting.org/beat/04X[illegible]11-10[illegible]CrimeWatch[illegible]API[illegible]“[illegible]Beat1X NCPC[illegible]……[illegible](OD)[illegible]……[illegible]”

Web

Ci ty of Oakland

2008 1

CTO Bob Glaze Ahsan Baig City Julian Ware

Andrew Wang Microsoft Excel

□□□

□□□□□□□□□□□□□□□□“□□□□”(te ideals of beauty)□□□□□□□□Crimespotting□□



□ 11-10□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□35□

□□

□□Feed□□□□□□□XML□API□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

□□□□□□□□□□□□Crimespotting□□□“□□□□□□”□□□□□FortiusOne□□□□Sean Go rman 2009□□□□□□□□□□□□□□□□□□□(htp□//blog.fortiusone.com/2009/01/28/data-is-the-public-good-data-is-the-infrastructure-data-is-the-stimulus/□□□“□□□□□”(posumer)□□□Web□□□

12 Sense.us

Jeffrey Heer

(Coes 2006 1940 14 3.5

“ ” “ ” AOL

—— ——

“ ” “ ”(lst mile) -

sense.us 150 (snse-making) Web

[illegible]——[illegible]——[illegible]
[illegible]
[illegible]

[illegible]

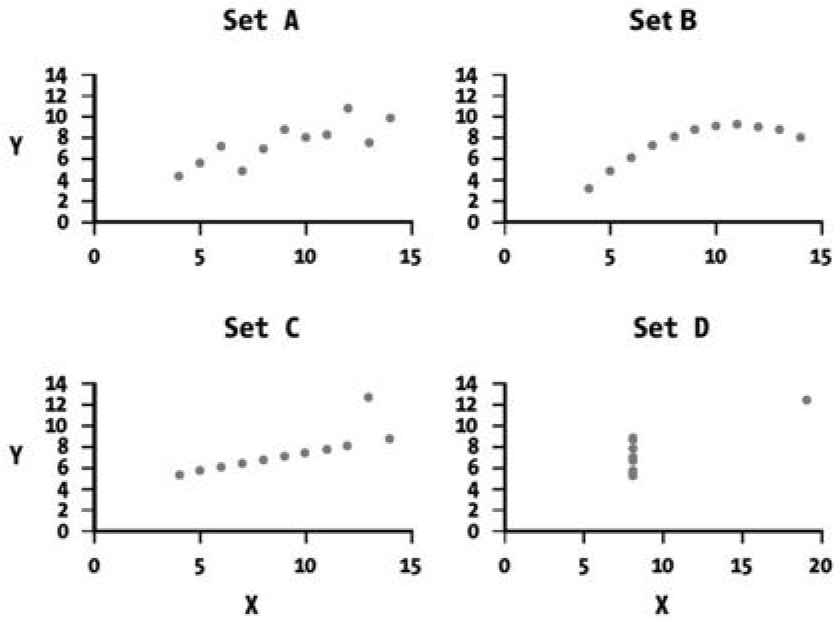
[illegible]70%[illegible]40%[illegible](Wre 2004[illegible]

[illegible]“[illegible]”[illegible](sns 3D[illegible]frills[illegible]

[illegible]Anscombe[illegible]“[illegible]”(Qartet)[illegible]Francis Anscombe[illegible](Ascombe 1973[illegible]12-1[illegible]

[illegible]Tufte[illegible]1997[illegible]Robison[illegible]2002[illegible]

[illegible]“[illegible]”(vsual variables)[illegible](Brtin 1967[illegible]Cart[illegible]1999[illegible]



Set A X	Set A Y	Set B X	Set B Y	Set C X	Set C Y	Set D X	Set D Y
10	8.04	10	9.14	10	7.46	8	6.58
8	6.95	8	8.14	8	6.77	8	5.76
13	7.58	13	8.74	13	12.74	8	7.71
9	8.81	9	8.77	9	7.11	8	8.84
11	8.33	11	9.26	11	7.81	8	8.47
14	9.96	14	8.1	14	8.84	8	7.04
6	7.24	6	6.13	6	6.08	8	5.25
4	4.26	4	3.1	4	5.39	19	12.5
12	10.84	12	9.11	12	8.15	8	5.56
7	4.82	7	7.26	7	6.42	8	7.91
5	5.68	5	4.74	5	5.73	8	6.89

Summary statistics

$\mu_X = 9.0$ $\sigma_X = 3.317$

$\mu_Y = 7.5$ $\sigma_Y = 2.03$

Linear regression

$Y = 3 + 0.5X$

$R^2 = 0.67$

[illegible] 12-1[illegible]Anscombe[illegible]“[illegible]”(Qartet)[illegible]

[illegible]

□□□□□□□□□□□□□□□□□□□□(snse-making)□□□□□□□

□□□

□□□□□□□□Martin Watternberg□□IBM□□□Fernanda Viégas□□□□□□□□□□□□□□□□□□“□□□□□□”□□□□□□□□□□□□□sense.us□□□□□□□□□□□□□□□□□□□□□□150□□□□□□□□□□□□□□□□□□□□□□□□□□□

□□□□□□□□□sense.us□□

[illegible]

Martin[illegible]Fernanda[illegible]

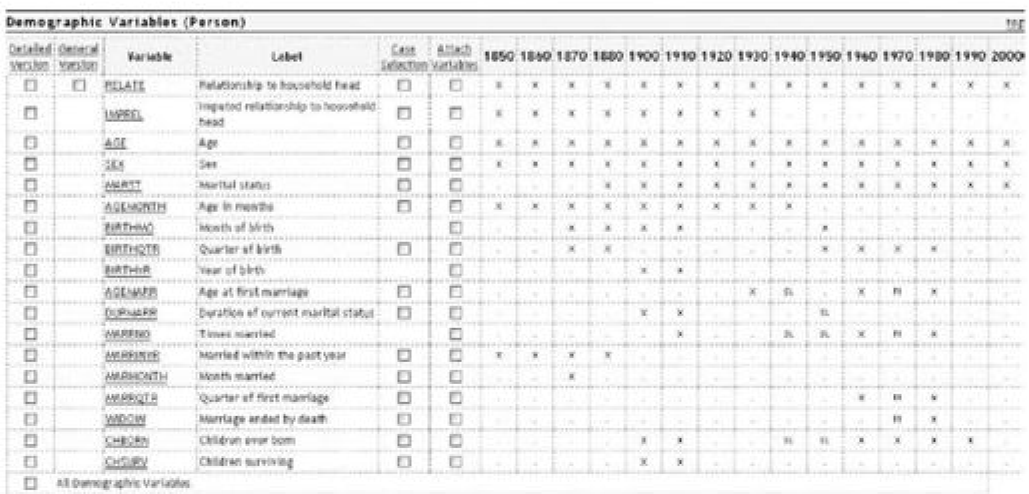
[illegible](htp[illegible]//www.census.gov)[illegible]

[illegible](htp[illegible]//www.ipums.org)[illegible](Itegrated Public Use Microdata Series,IPUMS)[illegible]IPUMS-USA[illegible]1850[illegible]2000[illegible]10[illegible]1890[illegible]1921[illegible](Cmmerce Building)[illegible]IPUMS[illegible]10[illegible]10[illegible]

1%[illegible]5%[illegible]

IPUMS-USA[illegible]IPUMS[illegible]“[illegible]”[illegible]

[illegible]IPUMS-USA[illegible]413[illegible]12-2[illegible]



Demographic Variables (Person)

Detailed Version	General Version	Variable	Label	Case Selection	Attach Variables	1850	1860	1870	1880	1900	1910	1920	1930	1940	1950	1960	1970	1980	1990	2000
☐	☐	RELATE	Relationship to household head	☐	☐	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
☐		IMPREL	Imputed relationship to household head	☐	☐	x	x	x	x	x	x	x	x	.	.	.	.	.	.	.
☐		AGE	Age	☐	☐	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
☐		SEX	Sex	☐	☐	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
☐		MARST	Marital status	☐	☐	.	.	.	x	x	x	x	x	x	x	x	x	x	x	x
☐		AGEMONTH	Age in months	☐	☐	x	x	x	x	x	x	x	x	x	.	.	.	.	.	.
☐		BIRTHMO	Month of birth		☐	.	.	x	x	x	x	.	.	.	x	.	.	.	.	.
☐		BIRTHQTR	Quarter of birth	☐	☐	.	.	x	x	.	.	.	.	.	x	x	x	x	.	.
☐		BIRTHYR	Year of birth		☐	.	.	.	.	x	x	.	.	.	.	.	.	.	.	.
☐		AGEMARR	Age at first marriage	☐	☐	.	.	.	.	.	.	.	x	SL	.	x	FH	x	.	.
☐		DURMARR	Duration of current marital status	☐	☐	.	.	.	.	x	x	.	.	.	SL	.	.	.	.	.
☐		MARRNO	Times married		☐	.	.	.	.	.	x	.	.	SL	SL	x	FH	x	.	.
☐		MARRINYR	Married within the past year	☐	☐	x	x	x	x	.	.	.	.	.	.	.	.	.	.	.
☐		MARRMONTH	Month married	☐	☐	.	.	x	.	.	.	.	.	.	.	.	.	.	.	.
☐		MARRQTR	Quarter of first marriage	☐	☐	.	.	.	.	.	.	.	.	.	.	x	FH	x	.	.
☐		WIDOW	Marriage ended by death	☐	☐	.	.	.	.	.	.	.	.	.	.	.	FH	x	.	.
☐		CHBORN	Children ever born	☐	☐	.	.	.	.	x	x	.	.	SL	SL	x	x	x	x	.
☐		CHSURV	Children surviving	☐	☐	.	.	.	.	x	x	.	.	.	.	.	.	.	.	.
☐	All Demographic Variables																			

[illegible] 12-2[illegible]IPUMS-USA[illegible]

IPUMS[illegible]“[illegible]”(ue it for good,never for evil)[illegible]

□□□□□IPUMS-USA□□sense.us□□□□□□□□□IPUMS Web□□□□□□□□□1850□2000□□□□□□□1890□□22□□□□□□□□□□□□□□□□□□□□□/□□□520MB□Gzip□□□□3.3GB□□□□

□□MySQL□□□□□□□□□□□□□□□□□□□(htp□//en.wikipedia.org/wiki/Star_schema)□□□□□□□□□□□□□□□□(fct table)□□□

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□“□□□”(rlled up)□□□

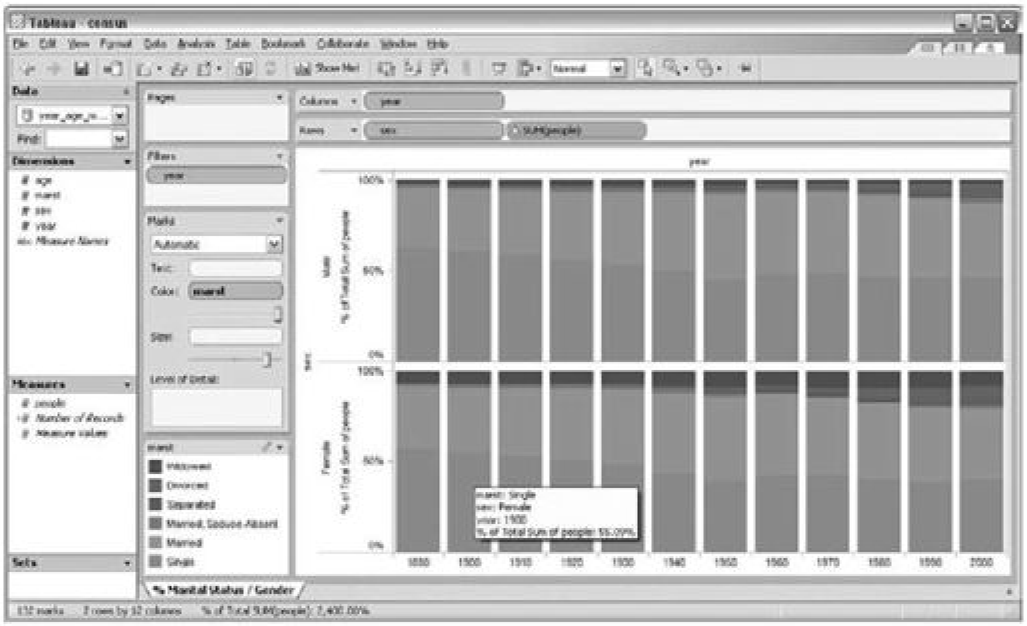
[illegible]

[illegible]

[illegible]

[illegible]SQL[illegible]Excel[illegible]Tableau[illegible]Tableau[illegible]Tableau[illegible]12-3[illegible]

[illegible]5[illegible]



[illegible] 12-3 [illegible] Tableau [illegible] 36 [illegible]

[illegible]

[illegible] Martin [illegible] Fernanda [illegible] sense.us [illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]200[illegible](sacked graph)[illegible]12-4[illegible]12-5[illegible]Martin[illegible](Bby Name Voyager)[illegible](Wttenberg[illegible]Kriss 2006[illegible]——[illegible]

[illegible]——[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

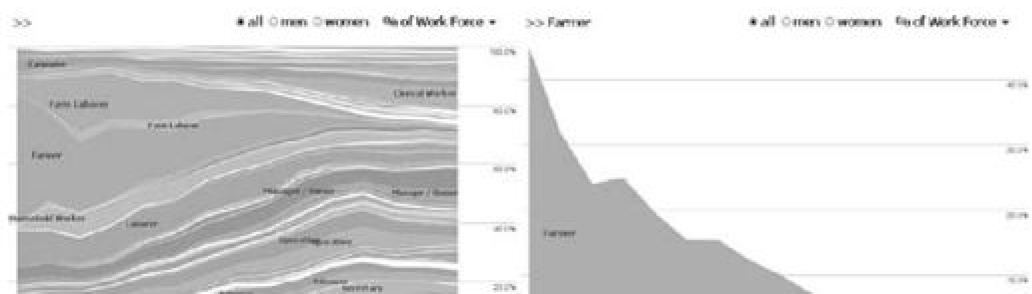
[illegible]“[illegible]”[illegible]1[illegible]

[illegible]

[illegible]

[illegible]

“[illegible]”(Jb Voyager)[illegible]150[illegible]12-4[illegible]



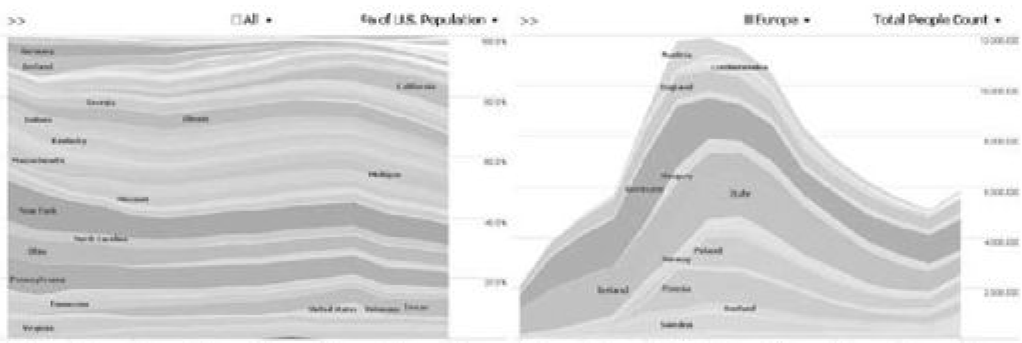
[illegible] 12-4[illegible]150[illegible]37[illegible]

[illegible]

Martin

“”(Brthplace Voyager)

12-5



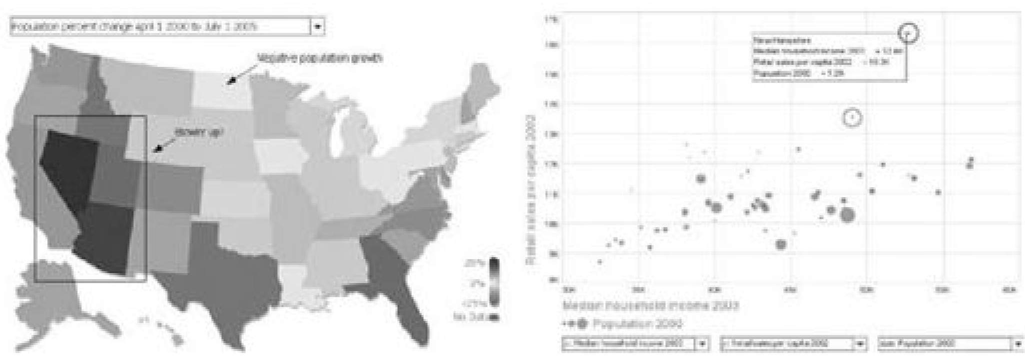
12-5 150

38

□□
□□□2000□
2005□□□□12-6□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□
□□□□North Dakota□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□
□□□□□□□□□□□□□□□□□□□□□□x□□□y□□□□□□□□□□□□□□□□□□□□□□□
□□□□□□□□□□□□□□□□□□□□□12-6□□□□□□□□□□□□□□□□□□□□□□□□□□
□□□□□□□□□□□□□□□□

□□□□□

□□12-
7□□□□□□□□□□□□□“□□-□□”□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□
□□□□□□□□□□□□□□□□□□□□□□□□□□□□y□□□□□□□□□5□□10□□□□□□□
□□□x□□12-
7□□



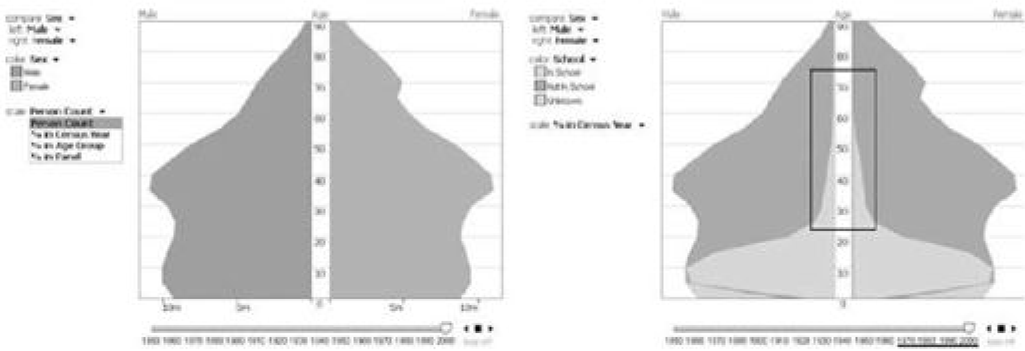
□ 12-6□□□□□□□□□□□□□□□□□□2000□2005□□□□□□□□□□□□□
□□□□□□□□□□□□x□□□□□□□y□□□□□□□□□□□□□□□□□□□□□□□□□39□

□□
□□

□□x,y□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

□□

□□(pevalence)□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□12-7□□□=□□□□□=□□□□□□□□□□□ColorBrewer(http□//colorbrewer.org)□□



□ 12-7□□□□□□□□□□□□□□□□2000□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□2000□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□40□□□

□□□“Georgia O'Keefe□□□□”(lva lamp)□

□□□□□□□□□□□□□□□□□□□□□□□□□□□□10□□□□□□□□□□□□□□□□□□□(pnel)□□

[illegible]

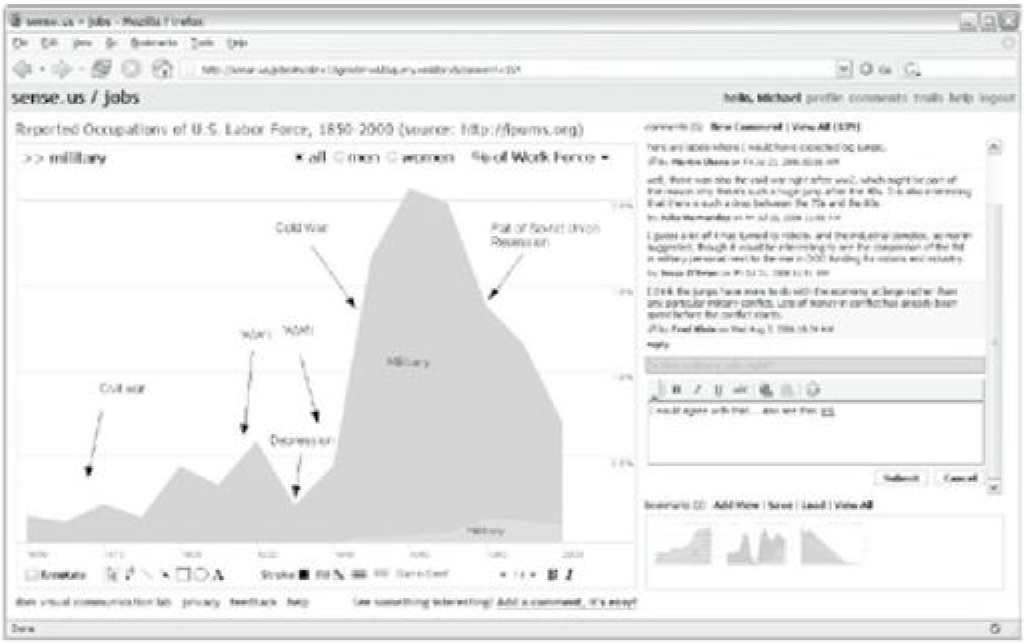
[illegible]Java Applet(Java[illegible]Web [illegible]Java[illegible]Flash[illegible]Java[illegible]Prefuse(http[illegible]//prefuse.org)[illegible]3GB[illegible]3MB[illegible]

[illegible]

[illegible]sense.us[illegible]12-8[illegible]Applet[illegible]12-8[illegible]a[illegible]12-8[illegible]b[illegible]12-8[illegible]c[illegible]12-8[illegible]d[illegible]e[illegible]

[illegible]

[illegible]sense.us[illegible]Web[illegible]URL[illegible]URL[illegible]URL[illegible]12-8[illegible]f[illegible]Email[illegible]



□　12-8□sense.us□□□□□□□□□a□□b□□□□□□□□□□□c□□□□□□□□□□□d□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□e□□□□□□□□□□□□□□□□f□□□□□□□□□□□URL□□URL□□□□□□□□□□□□□□□□41□

□□□□□□

□□□□□□□□□□□□□□□□□□□“□□□□□□”(dublylinked discussion)□□□Web□□□□□□□□□□□□□□□□□□□□□12-8□□e□□□

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□“□□□□”□“□□”□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□12-8□□d□□□

[illegible]

[illegible]“[illegible]”[illegible]

[illegible]12-8[illegible]“military”[illegible]12-8[illegible]f[illegible]“military”[illegible]“mili”[illegible]“mili”[illegible]——[illegible]URL[illegible]

[illegible]SHA-1[illegible]

[illegible]

[illegible]Web[illegible]12-8[illegible]b[illegible]Microsoft PowerPoint[illegible]

[illegible]12-8[illegible]e[illegible]

[illegible]XML[illegible]gzip[illegible]Base64[illegible]JavaScript[illegible]

[illegible]“[illegible]”[illegible]

[illegible]

[illegible]

[illegible]“[illegible]”(bokmark trail)[illegible]“[illegible]”[illegible]12-8[illegible]c[illegible]

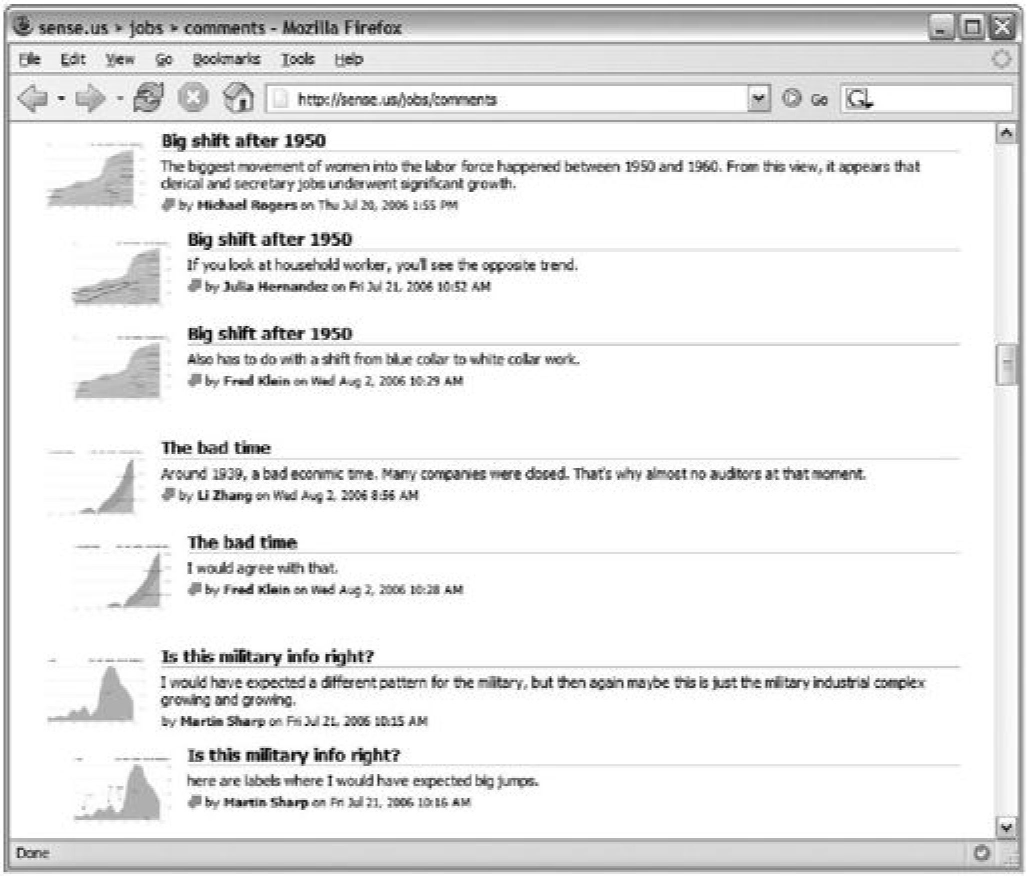
[illegible]

[illegible]

[illegible]“[illegible]”(Scial Navigation)[illegible]Web[illegible]

sense.us[illegible]12-9[illegible]

[illegible]



□　12-9□sense.us□□42□

□□□□□□

□□□“□□□□”□

“[illegible]”[illegible]“[illegible]”

[illegible]“[illegible]”[illegible]

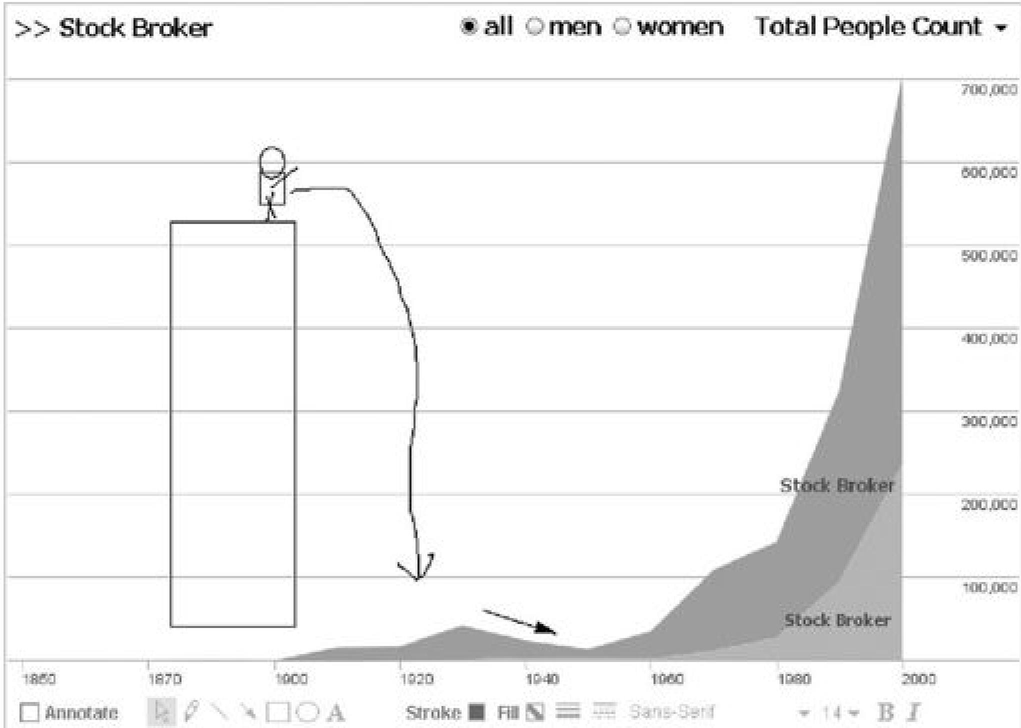
[illegible]30[illegible]sense.us[illegible]IBM[illegible]sense.us[illegible]

[illegible]

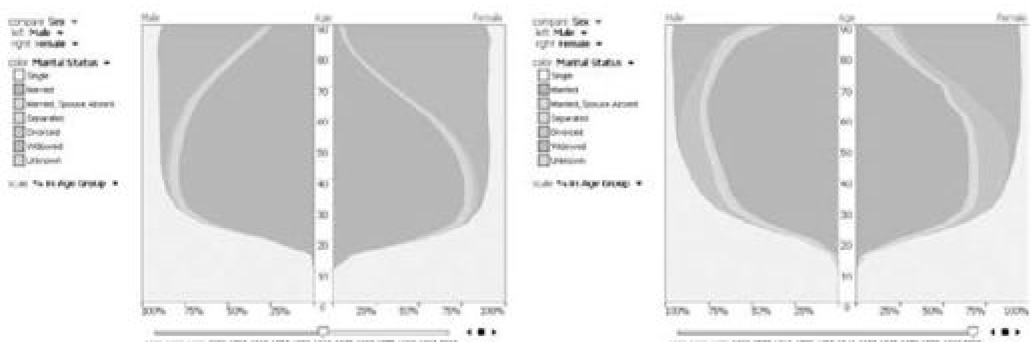
[illegible]“[illegible]”(savenger hunts)[illegible]

[illegible]1930[illegible]19[illegible]12-10[illegible]

[illegible]12-11[illegible]1960[illegible]1960[illegible]12-7[illegible]



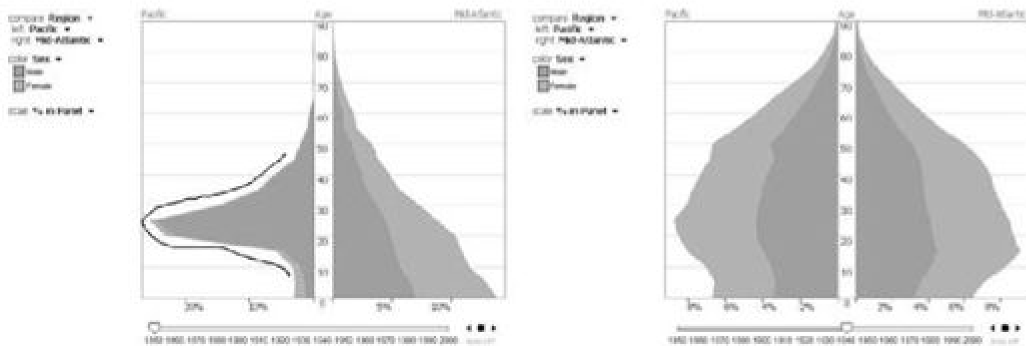
[illegible] 12-10[illegible]“[illegible]‘[illegible]’[illegible]”[illegible]43[illegible]



[illegible] 12-11[illegible]1940[illegible]2000[illegible]

[illegible]44[illegible]

[illegible]12-12[illegible]1850[illegible]“[illegible]”[illegible]90[illegible]10[illegible]



[illegible] 12-12[illegible]1850[illegible]1940[illegible]45[illegible]

[illegible]“[illegible]”[illegible]

[illegible]“[illegible]”[illegible]

[illegible]

[illegible]12-6[illegible]12-8[illegible]

[illegible]12-13[illegible]“[illegible]……[illegible]

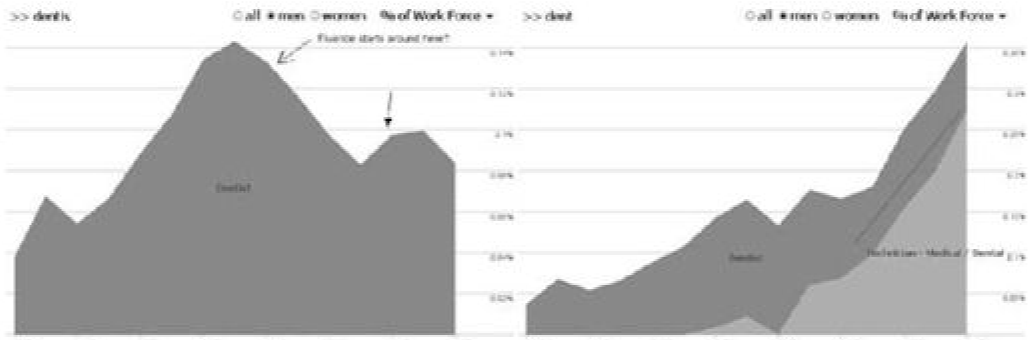
[illegible]”[illegible]“[illegible]

[illegible]”[illegible]

[illegible]“[illegible]

[illegible]”[illegible]

[illegible](snsemaking process)[illegible]



[illegible] 12-13[illegible]1930[illegible]46[illegible]

[illegible]IPUMS[illegible]20[illegible]50[illegible]

[illegible]“[illegible]”(Oerative)[illegible]“[illegible]”[illegible]“[illegible]”[illegible]

[illegible]“[illegible]”[illegible]“[illegible]”(sgnposts)[illegible]16%[illegible]

[illegible]“[illegible]”

[illegible]“[illegible]”[illegible]“[illegible]”[illegible]“[illegible]”[illegible]

[illegible]“[illegible]”[illegible]“[illegible]”[illegible]“[illegible]”[illegible]

[illegible]

[illegible]sense.us[illegible]sense.us[illegible]IBM[illegible]sense.us[illegible]Many-Eyes.com[illegible]wiki[illegible]

[illegible]Web[illegible]Swivel.com[illegible]Data360.org[illegible]Spotfire Decision Site Posters[illegible]Tableau Server[illegible]Web[illegible]Web[illegible]

[illegible]sense.us[illegible]sense.us[illegible]

[illegible]

参考文献

[1]Anscombe,Francis J.（1973）."Graphs in Statistical Analysis."American Statistician，27，17-21.

[2]Bertin,Jacques.（1967）. Sémiologie Graphique,Gauthier-Villars.English translation by W.J Berg as Semiology of Graphics,University of Wisconsin Press，1983.

[3]Card,Stuart K.，Ben Shneiderman,and Jock D. Mackinlay.（1999）.Readings in Information Visualization:Using Vision To Think,Morgan-Kaufmann.

[4]Cleveland,William S. and Robert McGill.（1985）."Graphical Perception and Graphical Methods for Analyzing Scientific Data."Science，229（4716），828-833.

[5]Croes,Marnix.（2006）."The Holocaust in the Netherlands and the Rate of Jewish Survival."Holocaust and Genocide Studies，20（3），474-499.

[6]Robison,Wade,Roger Boisjoly,David Hoeker,and Stefan Young.（2002）."Representation and Misrepresentation:Tufte and the Morton Thiokol Engineers on the Challenger."Science and Engineering Ethics，8，59-81.

[7]Tufte,Edward R.□1997□. Visual Explanations:Images and Quantities,Evidence and Narrative,Graphics Press.

[8]Ware,Colin.□2004□. Information Visualization:Perception for Design,Second Edition,Morgan-Kaufmann.

[9]Wattenberg,Martin and Jesse Kriss.□2006□."Designing for Social Data Analysis."IEEE Transactions on Visualization and Computer Graphics□12□4□□549-557.

[illegible]13[illegible] [illegible]

Coco Krumme

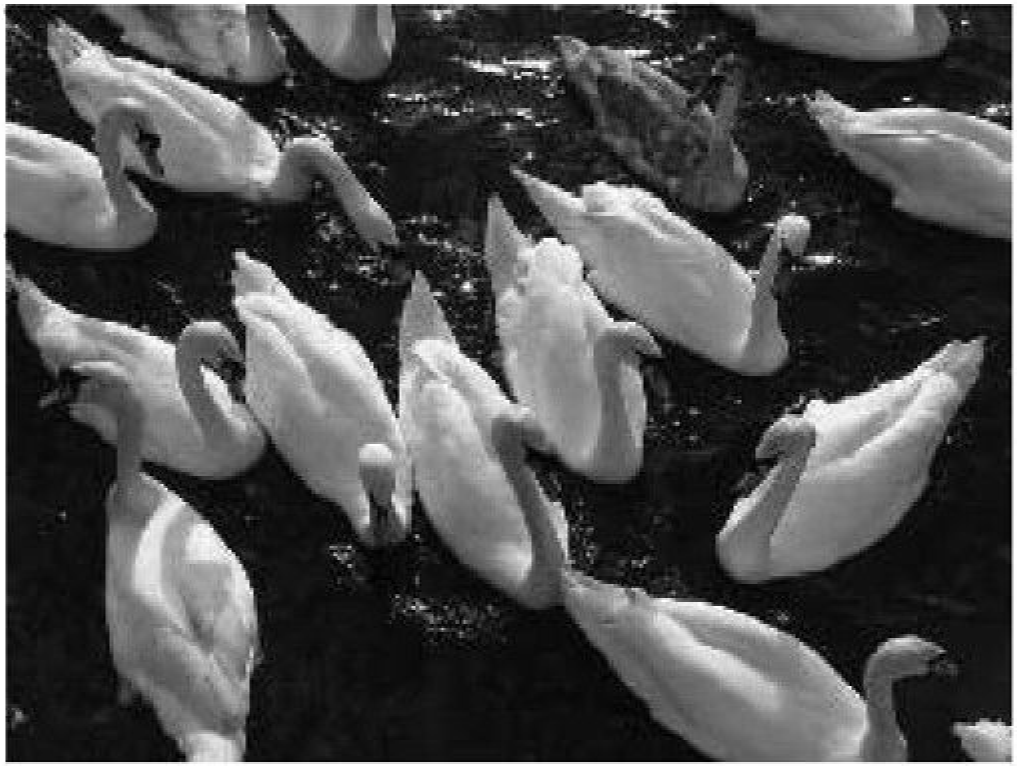
[illegible]

[illegible]20[illegible]“[illegible]”[illegible]

[illegible]

[illegible]13-1[illegible]

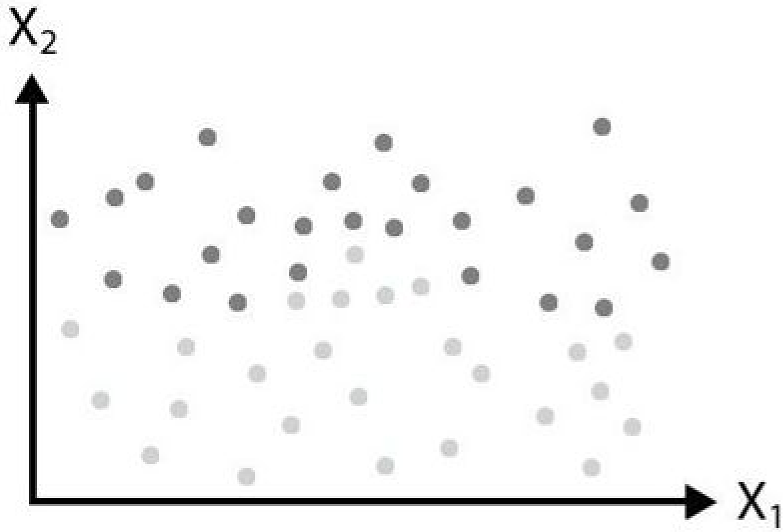
[illegible]



[illegible] 13-1[illegible]

[illegible]

[illegible]13-2[illegible]

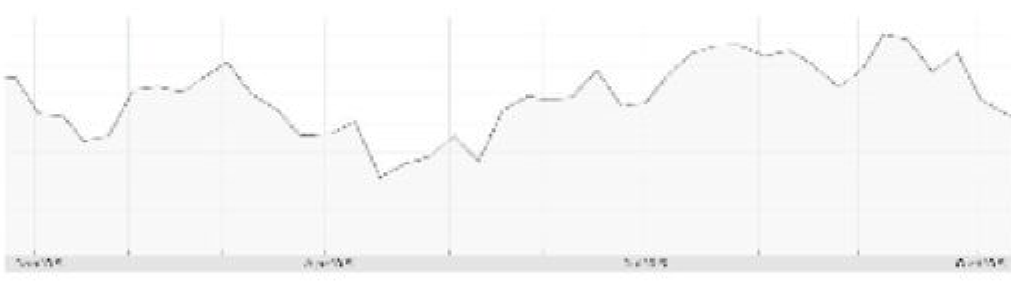


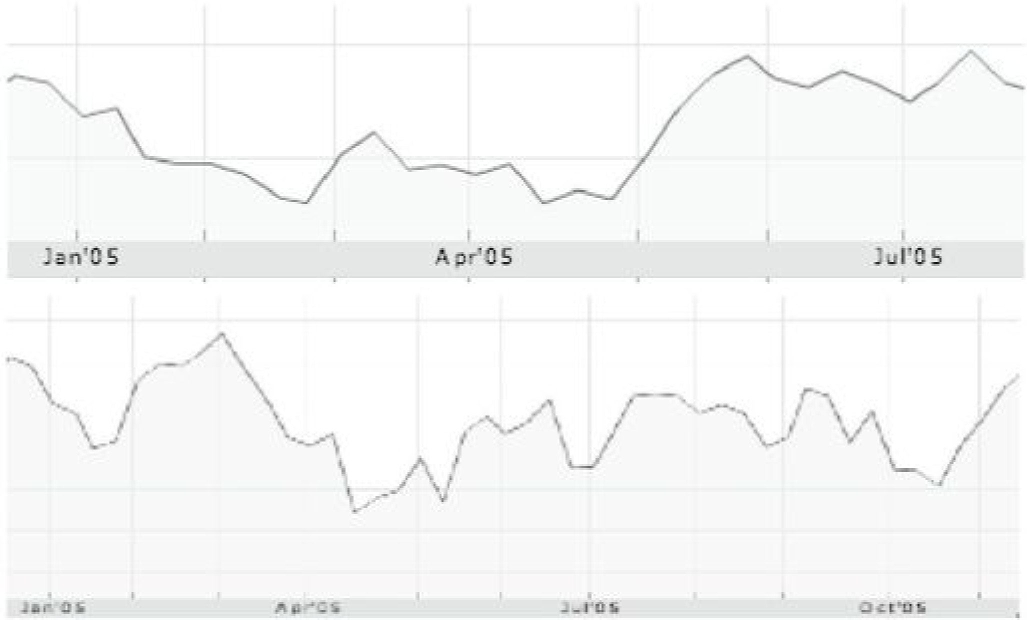
[illegible] 13-2[illegible]47[illegible]

[illegible]

[illegible]13-3[illegible]2005[illegible]1[illegible]10[illegible]2006[illegible]

[illegible]——[illegible]10[illegible]





□ 13-3□□□□□□a□b□c□□2005□□□□□□□□48□

□□□□□□□□□□□□□□□□□“□□□□”(nrrative fallacy)□□□□□□□□□□□

□□□□□□□□□□□□□□□□□□□□MBA□□——□□□□□□□□□□□□——□□□□□□

□□□□□□□□□□□“□□”□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

□“□□□□”□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□——□□□□□□□□□□

□□□□□□□□□□□□□□□(Kumme□□□□□□□□□□□□□□□□□□□□□□□“□□□□□

□□□□□□□□□□□□□100□□”□

□□

□□

□□□□

□□“□□□

□”(cnfirmation bias)(Lrd□1979□□□□□□□□□□□□□□□□□□□□□□□□□

□□

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□(Jng 2006□□□□□□□□□□□

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□(L□Repin

2002□□

□□

□□

·□□A□100%□□7400□□

·□□B□75%□□□□10000□□□25%□□□□□□

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

□□□□□□□□□□□□□□□□□□□□□□

·□□C□100%□□7400□□

·□□D□75%□□□□10000□□□25%□□□□

□□□□□□□A□D□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

□□□

□□□

B+C□□□□□□□□□□□□□□□1000□□□□A+D□□□□□□□□□□□-1000□□□□

□□□□□□A+D□□□□□B+C□□□□□□□□□□□2000□□□□

Daniel Kahneman□Amos Tversky□20□□70□□□□□□□□□□□□□

□□

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

[illegible]

[illegible]“[illegible]”[illegible]

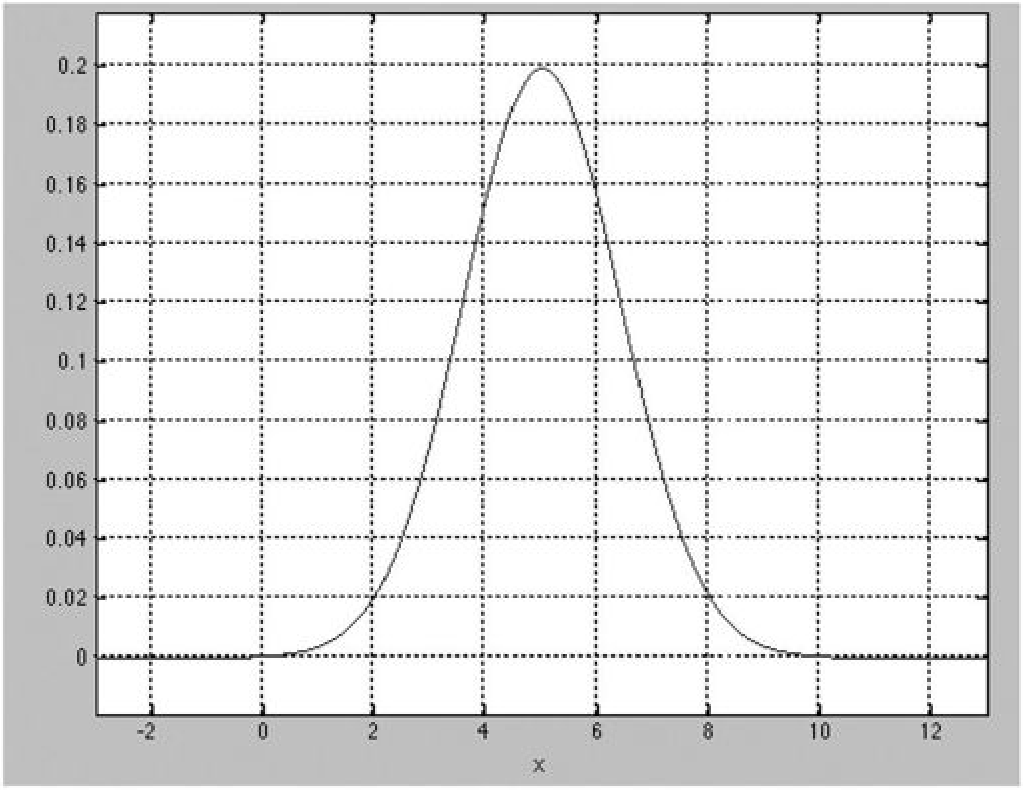
·[illegible]

·[illegible]

[illegible]“[illegible]”[illegible]

[illegible]

[illegible]13-4[illegible]



[illegible] 13-4[illegible]49[illegible]

[illegible]

[illegible]Karl Popper[illegible]

[illegible]

[illegible]“[illegible]”(flse positive)[illegible]“[illegible]”(flse negative)[illegible]

□□

□“□□□”□□□□□□□□□□□□□□□□□□□□□“□□□”□□□□

□□□□□□□□□

□□□□□□□□□□□□□□□□□□□□□□“□□”□□□□10TB□□□□□□10□□□□□□□

□□□□□□100□□□□□□□□□□□10□□□□□

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□X□□□□□□□□□□□□□

□□□□□□□□□□“□□□□”(Rd Queen)□□[1]□□□□□□□□□□□□□□□□□□□□□

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

□□□□□□□□□□□□□□□□□□□□□□□□□□□“401(k)□□”□401(k)plans)

[2]□□□

□□□□□□□□□□□□□□□□(Ienger□Lepper 2000□□

□□

□□□□□□□□□□□□□□□□□□□□□□□□□□□□——□□□□□□□□□□□□□□□□□□□

□□□□

□□□□□□□□

□□□□□□□□□□□□□□□□□□□□□(crrelation)□□□□□(cusality)□“□□□

□”(srange bedfellows)□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

□□

□□□□□□□“□□”□□□□□□□□□□□□□□□□□□□□□□□□□□□

(sellenbarger 2008□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

[illegible]“[illegible]”[illegible]

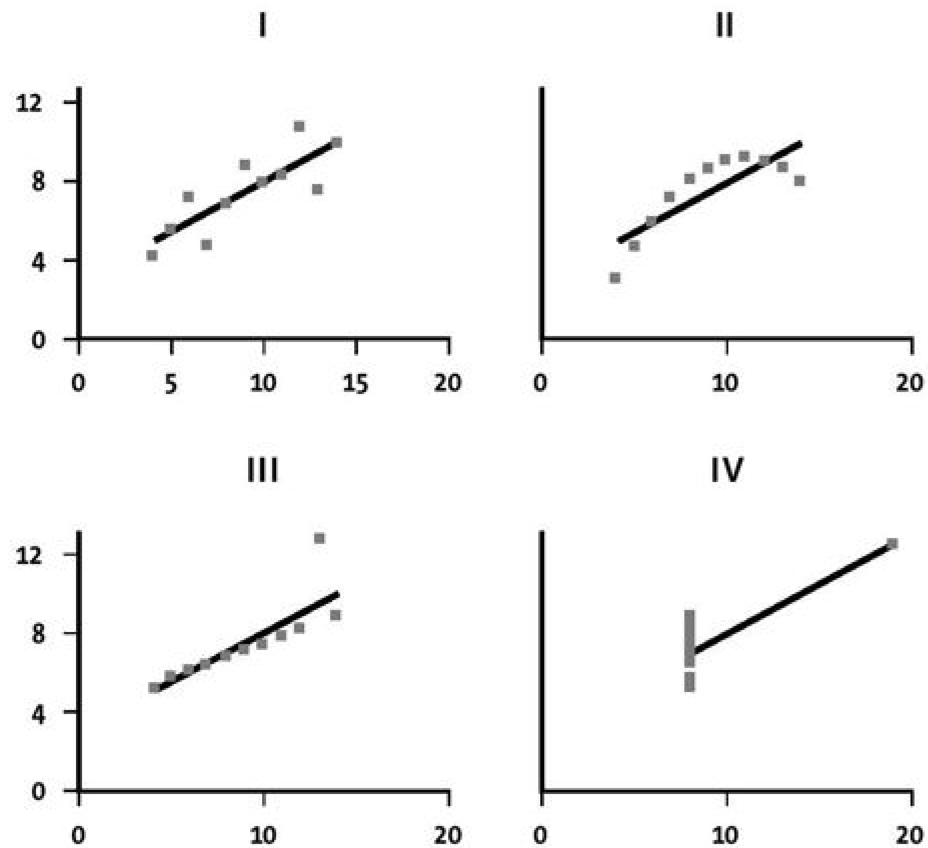
[illegible]

[illegible]

[illegible]

[illegible]

[illegible]13-5[illegible]——[illegible]——[illegible](Ascombe 1973[illegible]

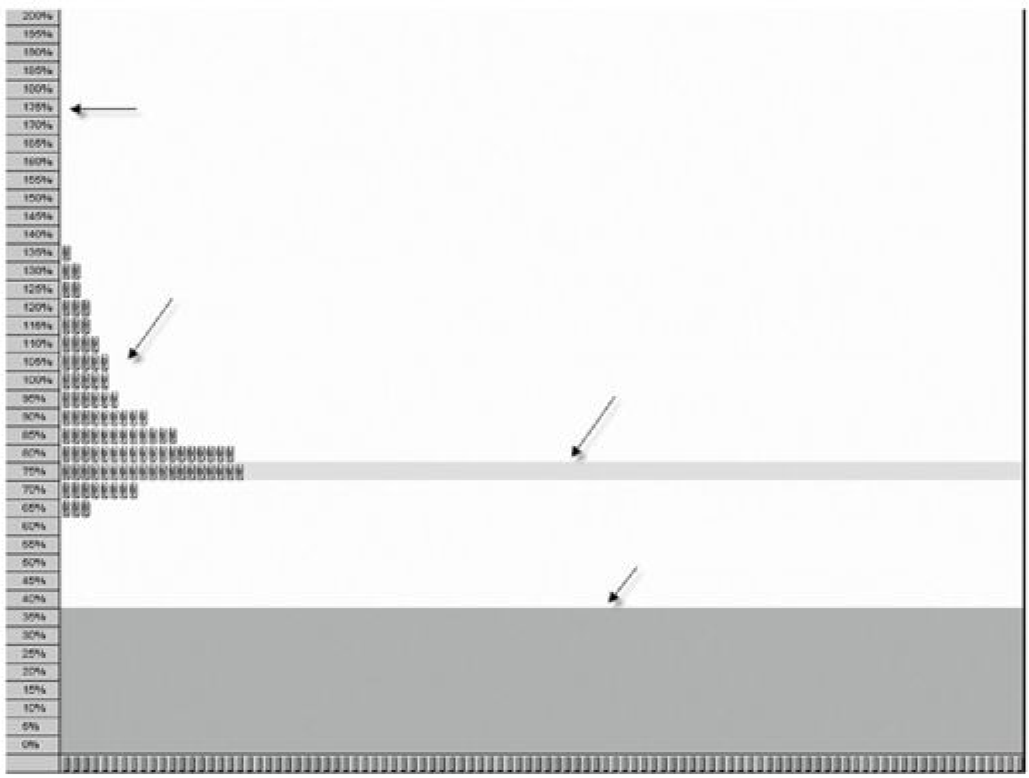


□　13-5□Anscombe□“□□□”(qartet)□□□□□□□□□□□□□□□□

　　□□□
□□□——□□□□□□□□FDA□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□
□□□□□□□□□□□□□□□□□□□□□5□□□□□□□□□□□□——□□□□□□□□□□□□□
□□□□□□□□□□□□□□□□

　　□□□
□□——□□
□□□□□□□□□□□□□□□——□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□
□□
□□□□□□□□□□□□□□□□□□□□□□□□“□□□□□”(panning fallacy)□□

□□13-6□□□□□□□□□100□“□□□□”□□□□□□□□□□□□□□□□□□□□□□□□□75%□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□“□□”(g)□□□□□□□□□□□□□□□□□□□□□□□“□□”(otcome)(Gldstein□2008□□□□□□□□□□□□□□□□□□□□□□□□□□□□100□□□□□□□□□□□□□



□ 13-6□Goldstein□□□□□□□□□□□□□□□□□□□□□□□50□

□□□□Stephen Jay Gould□□□□□□□□□□□□□□□□□□□□□“□□□□□□□□□□□”(Te Median is not the Message)□Gould□□□□□□“□□□8□□”□□□□□□□□□□□□□□□“□□□□□□□□”□□□□□□□□——□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□“8□□”□□□□□□□□□□□□□Gould□□□□□□□□□□□□

[illegible]

[illegible]

[illegible]2012[illegible]500[illegible]

[illegible]——[illegible]

[illegible]9.8[illegible]Popper[illegible]“[illegible]”(flsifiability)[illegible]

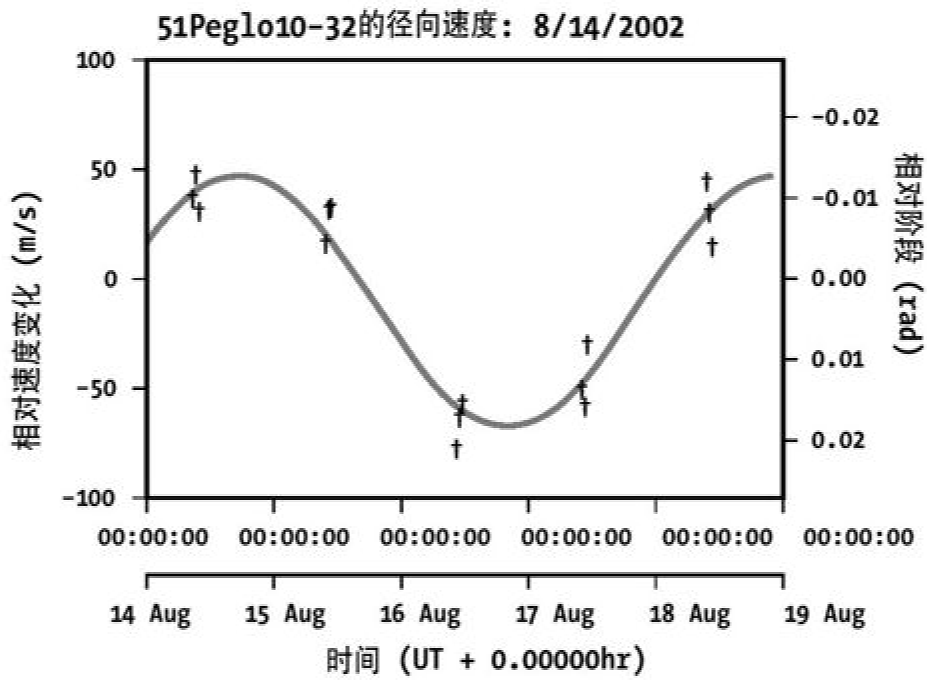
[illegible]

[illegible]15[illegible]13-7[illegible](G[illegible]2004[illegible]

[illegible]

[illegible]200[illegible]18[illegible]20[illegible]GPS[illegible]21[illegible]20[illegible]

[illegible]1914[illegible]1975[illegible]“[illegible]”[illegible]



[illegible] 13-7[illegible]

[illegible]“[illegible]”[illegible]2005[illegible]

[illegible]……[illegible]

[illegible]

[illegible]Joshua Epstein 2008[illegible]“[illegible]”

[illegible]

[illegible]“[illegible]”(fogging horse)[illegible]

[illegible]100[illegible]1000[illegible]0.1%[illegible]1%[illegible]

[illegible]

[illegible]99%[illegible]1%[illegible]9.9%[illegible]99.9[illegible]9990[illegible]990[illegible]990/9990[illegible]9.9%[illegible]

[illegible]

[illegible]Daniel Kahneman——[illegible]——[illegible]“[illegible]”(bse rate fallacy)[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]Fisher[illegible]“[illegible]z[illegible]”[illegible]z[illegible]Fisher[illegible]p=0.01[illegible]p=0.06[illegible]p=0.05[illegible]

[illegible]Fisherian[illegible]

[illegible]——[illegible](Bkhchandani[illegible]1998[illegible]

[illegible]

[illegible]“[illegible]”(mtrix file)[illegible]

[illegible]P2P(peer-to-peer)[illegible]Prosper.com[illegible]35[illegible]75%[illegible]——[illegible]100[illegible]——[illegible]

[illegible]

[illegible]“[illegible]”[illegible]

[illegible]Amazon[illegible]Mechanical Turk)[illegible]Prosper.com[illegible]

[illegible]——[illegible]——[illegible]“[illegible]”[illegible]

[illegible]“[illegible]”

[illegible]“[illegible]”[illegible]

[illegible]Proper[illegible]API[illegible]

[illegible]

[illegible]Fisher[illegible]Bayes[illegible]

[illegible]“[illegible]”[illegible]“Lake Wobegon[illegible]”[illegible]

[illegible]

[1]Lewis Carroll[illegible]“[illegible]”[illegible]“[illegible]”[illegible]

[2]401(k)[illegible]20[illegible]80[illegible]401[illegible]k[illegible]

[illegible]

[illegible]——[illegible]——[illegible]

[illegible]

□□□□

[1]Anscombe,F. J."Graphs in Statistical Analysis."The American Statistician,vol.27□no.1(Fbruary 1973□□pp.17-21.

[2]Bikhchandani,S. et al.□1998□."Learning from the Behavior of Others:Conformity,Fads,and Informational Cascades□"Journal of Economic Perspectives,vol.12□issue 3□pp.151-170.

[3]Dwyer,D. W.2005."Examples of overfitting encountered when building private firm default prediction models" (ww.moodyskmv.com/research/files/wp/Overfitting_Private_Firm_Models.pdf).

[4]Epstein,Joshua M.□2008□."Why Model?". Journal of Artificial Societies and Social Simulation 11□4□12.

[5]Ge,J. et al.□2004□"All Sky Extrasolar Planet Searches with Multi-Object Dispersed Fixeddelay Interferometer in Optical and near-IR."Proc.SPIE□5492□711.

[6]Goldstein,Daniel G. et al.□2008□."Choosing Outcomes Versus Choosing Products:Consumer-Focused Retirement

Investment Advice."Journal of Consumer Research，35(Otober)，440-456.

[7]Gould,S. J."The Median is not the Message."Discover Magazine，1985.

[8]Iyengar,S. S.and M.R.Lepper."When choice is demotivating:can one desire too much of a good thing?"Journal of Personality and Social Psychology，2000，vol.79，no.6，995-1006.

[9]Jeng,M."A selected history of expectation bias in physics."American Journal of Physics，2006.

[10]Kahneman,Daniel and Amos Tversky（1979）."Prospect Theory:An Analysis of Decision under Risk，"Econometrica,XLVII（1979），263-291.

[11]Krumme,C."Telling tales:the effects of narrative creation on decision-making with data，"working paper.

[12]LO'A. W.and D.V.Repin."The Psychophysiology of Real-Time Financial Risk Processing"Journal of Cognitive Neuroscience,April 1，2002，vol.14，no.3，323-339.

[13]Lord,C. G.et al.（1979）."Biased assimilation and attitude polarization:The effects of prior theories on

[14]subsequently considered evidence."Journal of Personality and Social Psychology□37□2098-2109.

Shellenbarger,Sue.Wall Street Journal"Work and Family"column,March 22□2008.

[illegible]14[illegible] [illegible]

Peter Norvig

[illegible]“[illegible](Budelaire)[illegible]”[illegible]“[illegible]”[illegible]“[illegible](Toreau)[illegible]”[illegible]“[illegible]”[illegible]“[illegible]”[illegible]Web[illegible]1MB[illegible]Web[illegible]“[illegible]”[illegible](Lugh Out Loud,LOL)[illegible](Rckrolling)[illegible]·[illegible]

G[illegible]Thorsten Brants[illegible]Alex Franz[illegible]2006[illegible]1MB[illegible](Lnguistic Data Consortium)[illegible](htp[illegible]//tinyurl.com/ngrams)[illegible](to-word)[illegible]“the”[illegible]230[illegible]1MB[illegible]2.2%[illegible]“rebating”[illegible]12750[illegible]“fnuny”[illegible]“funny”[illegible]“Find all posts”[illegible]1300[illegible]0.001%[illegible]“each of the”[illegible]1[illegible]“All Rights Reserved”[illegible]0.01%[illegible]

```
outraged many African 63
outraged many Americans 203
outraged many Christians 56
outraged many Iraqis 58
outraged many Muslims 74
outraged many Pakistanis 124
outraged many Republicans 50
outraged many Turks 390
outraged many by 86
outraged many in 685
```

outraged many liberal 67
outraged many local 44
outraged many members 61
outraged many of 489
outraged many people 444
outraged many scientists 90

[illegible]

[illegible]——[illegible]

[illegible](tlk the talk)——[illegible]token[illegible]——[illegible]token[illegible](tpe)[illegible]“Run,Lola Run”[illegible]token[illegible]token)[illegible]G[illegible]1MB[illegible]token[illegible]1300[illegible]100[illegible]“www.njstatelib.org”[illegible]“+170.002”[illegible]“1.5GHz/512MB/60GB”[illegible]“Abrahamovich”[illegible]10[illegible]1/3[illegible]token[illegible]1000[illegible]2/3[illegible]10[illegible]98%[illegible]

1-token[illegible]2-token[illegible]n-token[illegible]N[illegible]P[illegible]P(the)=.022[illegible]token“the”[illegible]0.022[illegible]2.2%[illegible]W[illegible]token[illegible]W3[illegible]token[illegible]W1[illegible]3[illegible]token[illegible]P(Wi=the|Wi-1=of)[illegible]token“the”[illegible]token“of”[illegible]

G□□□□□□□□□□□□□□□□□200□□□□□□□□□□□□□□□UNK□□□□□□□□

□□□□40□□N□□□□□□□□□□□□□□□□□□□□□□□□□24G□□□□□□□□□□□□□

□□□□□□□□□□□□S□□□□□/S□□□□

□□□□□□□□□□□□□□□□□□□□□□

[illegible]

[illegible]“[illegible]”[illegible]“float like a butterfly”[illegible]5[illegible]URL[illegible]

[illegible]“choosespain.com”[illegible]“chooses pain”[illegible]“choose Spain”[illegible]3210[illegible]“chooses pain”[illegible]1MB[illegible]40[illegible]“choose Spain”[illegible]80[illegible]

[illegible]“insufficientnumbers”[illegible]

```
insufficient numbers 20751
in sufficient numbers 32378
```

“In sufficient numbers”[illegible]“insufficient numbers”[illegible]

1.[illegible]“choose Spain”[illegible]“choose Spain”[illegible]n[illegible]——[illegible]——[illegible]

2.[illegible]“insufficient numbers”[illegible]“in sufficient numbers”[illegible]“in suffi cient numb ers”[illegible]“hello world”[illegible]——[illegible]

3.[illegible]

[illegible]

best=argmaxc∈candidates P(c)

[illegible]

```
best=max(c in candidates,key=P）
```

[illegible]

```
>>> segment（'choosespain'）
['choose'，'spain']
```

[illegible]1——[illegible]

P(W1[illegible]n[illegible]=Πk=1[illegible]nP(Wk|W1[illegible]k-1[illegible]

[illegible]N[illegible]

[illegible]30G[illegible]0[illegible](bcking off)[illegible]

P(W1[illegible]n[illegible]=Πk=1[illegible]nP(Wk)

[illegible]'wheninrome'[illegible]when in rome[illegible]P(when)×P(in)×P(rome)[illegible]

n[illegible]2n-1[illegible]n-1[illegible]

[illegible]'wheninthecourseofhumaneventsitbecomesnecessary'[illegible]35T[illegible]“w”[illegible]“wh”[illegible]“whe”[illegible]“when”[illegible]

[illegible]segment[illegible]L=20[illegible]

□□□□□□□P(first)×P(remaining)□□□□□□□□□□□

□□“when”□□□□□□□□□□□□□□□□□5□□□

first	P(*first*)	P(*remaining*)	P(*first*) × P(*remaining*)
w	$2\cdot10^{-4}$	$2\cdot10^{-33}$	$6\cdot10^{-37}$
wh	$5\cdot10^{-6}$	$6\cdot10^{-33}$	$3\cdot10^{-38}$
whe	$3\cdot10^{-7}$	$3\cdot10^{-32}$	$7\cdot10^{-39}$
when	$6\cdot10^{-4}$	$7\cdot10^{-29}$	$4\cdot10^{-32}$
wheni	$1\cdot10^{-16}$	$3\cdot10^{-30}$	$3\cdot10^{-46}$
whenin	$1\cdot10^{-17}$	$8\cdot10^{-27}$	$8\cdot10^{-44}$

□□□□□□□□Python□□□□□□□□

```
@memo
def segment(text)□
"Return a list of words that is the best segmentation of text."
if not text:return[]
candidates=□[first]+segment(rem)for first,rem in splits(text)□
return max(candidates,key=Pwords)
def splits(text,L=20□□
"Return a list of all possible(first,rem)pairs,len(first)□=L."
return[(txt[□i+1]□text[i+1□]□
for i in range(min(len(text)□L□□]
def Pwords(words)□
"The Naive Bayes probability of a sequence of words."
return product(Pw(w)for win words)
```

□□□□□□——□□□□□□□□□□□□product□□□□□□□□□□□□□□□□memo□□□□□□(dcorator)□□□□product□□□□□□□□□□□□□□□□□□□□□□□Pw□□□□□□□□□□□□□□□□□□□□□

□□□□□memo□□□□□□□n□□□□□□□□□□□□□□□□□2n□□□□□□□□memo□□□□□□n□□□——memo□□□□□□□□□□□□□□□□□□□□n□□□□□

□□□□□□O(L)□□□□□□□□□□□□O(n)□□□□□□□□□□□□□□□□□□□□□□□

O(n^2L□□

□□P w□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

Count(word)/N□□□N□□□□□□□□□□□□□□□□□1300□□□□□□□□□□□□□

□□□vocab_common□□□□□□□□□□1□□□□□□□□□

□“the”□“The”□“THE”□□□□□□□□□□□□□□□□□“the”□□2□□□□□□□□□

□□□□□□□□□□□□□□□□“+170.002”□□□□□□□□□“can't”□□□□□3□

100□□□□□□□□1/3□□□

Pw□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□10000□□

□□□□□□□□□□□□□□□□□□□□□□□□□□0□□□□□□□□□□□□□□□□□□□□□□

token□□N□□□□10000□□vocab_common□□□□□□□□□□12711□□□

□□□□□□□□□□□□□□□□□0□12711/N□□□□□□□□□□□□□□□□□□□□□□□

20□□□□□□□□□□□□□□□□□□□□6□□□□□□□□□□□□□□□□□□□□□□Pdist□

□□□□□□□□(ky,count)□□□□□□□□□□□□□□□□□□□□□□□1/N□□□Pdist□

□□□□□□□□□□□□□□□□□□□□□□□(oerride)□□□□□□□□□□□□□□□□□□□□

□□□□□□□□□□□□□□□□10/N□□□□□□□□□□□□□□□□□10□□□□□□□□Pw□

□□□□□Pdist□

```
class Pdist(dict)□
"A probability distribution estimated from counts in datafile."
def__init__(slf,data,N=None,missingfn=None)□
for key,count in data□
self[key]=self.get(key□0□+int(count)
self.N=float(N or sum(self.itervalues□□□□
self.missingfn=missingfn or(lambda k,N□1./N□
def__call__(slf,key)□
```

```
if key in self:return self[key]/self.N
else:return self.missingfn(key,self.N□
def datafile(name,sep='\t'□□
"Read key,value pairs from file."
for line in file(name)□
yield line.split(sep)
def avoid_long_words(word,N□□
"Estimate the probability of an unknown word."
return 10./□N*10**len(word)□
N=1024908267229##Number of tokens in corpus
Pw=Pdist(datafile□'vocab_common'□□N,avoid_long_words)□
```

□□□Pw[w]□□□w□□□□□□□Pw(w)□□□□□□□□□□□□□□□□□□□□□□

http□//norvig.com/ngrams□□□

□□□□□□□□□□□□□□□□□□□□□□□□□□

```
□□□segment□'choosespain'□
['choose'□'spain']
□□□segment□'thisisatest'□
['this'□'is'□'a'□'test']
□□□segment□'wheninthecourseofhumaneventsitbecomesnecessary'□
['when'□'in'□'the'□'course'□'of'□'human'□'events'□'it'□'becomes'□'necessary']
□□□segment□'whorepresents'□
['who'□'represents']
□□□segment□'expertsexchange'□
['experts'□'exchange']
□□□segment□'speedofart'□
['speed'□'of'□'art']
□□□segment□'nowisthetimeforallgood'□
['now'□'is'□'the'□'time'□'for'□'all'□'good']
□□□segment□'itisatruthuniversallyacknowledged'□
['it'□'is'□'a'□'truth'□'universally'□'acknowledged']
□□□segment□'itwasabrightcolddayinaprilandtheclockswerestrikingthirteen'□
['it'□'was'□'a'□'bright'□'cold'□'day'□'in'□'april'□'and'□'the'□'clocks'□
'were'□'striking'□'thirteen']
□□□segment
□'itwasthebestoftimesitwastheworstoftimesitwastheageofwisdomitwastheage
offoolishness'□
['it'□'was'□'the'□'best'□'of'□'times'□'it'□'was'□'the'□'worst'□'of'□'times'□
'it'□'was'□'the'□'age'□'of'□'wisdom'□'it'□'was'□'the'□'age'□'of'□
'foolishness']
□□□segment
□'asgregorsamsaawokeonemorningfromuneasydreamshefoundhimselftransform
```

ed
inhisbedintoagiganticinsect'□
['as'□'gregor'□'samsa'□'awoke'□'one'□'morning'□'from'□'uneasy'□'dreams'□'he'
□
'found'□'himself'□'transformed'□'in'□'his'□'bed'□'into'□'a'□'gigantic'□
'insect']
□□□segment
□'inaholeinthegroundtherelivedahobbitnotanastydirtywetholefilledwiththe
endsofwormsandanoozysmellnoryetadrybaresandyholewithnothinginittositdo
wnonortoeat
itwasahobbitholeandthatmeanscomfort'□
['in'□'a'□'hole'□'in'□'the'□'ground'□'there'□'lived'□'a'□'hobbit'□'not'□'a'□
'nasty'□'dirty'□'wet'□'hole'□'filled'□'with'□'the'□'ends'□'of'□'worms'□'and'□
'an'□'oozy'□'smell'□'nor'□'yet'□'a'□'dry'□'bare'□'sandy'□'hole'□'with'□
'nothing'□'in'□'it'□'to'□'sitdown'□'on'□'or'□'to'□'eat'□'it'□'was'□'a'□
'hobbit'□'hole'□'and'□'that'□'means'□'comfort']
□□□segment
□'faroutintheunchartedbackwatersoftheunfashionableendofthewesternspiral
armofthegalaxyliesasmallunregardedyellowsun'□
['far'□'out'□'in'□'the'□'uncharted'□'backwaters'□'of'□'the'□'unfashionable'□
'end'□'of'□'the'□'western'□'spiral'□'arm'□'of'□'the'□'galaxy'□'lies'□'a'□
'small'□'un'□'regarded'□'yellow'□'sun']

□□□□□□□□□□□□□□□□□□□“Samsa”□“oozy”□□□□□□□□□□□□□□□□□□“Samsa”□10000□□□□□□□□□□42000□□“oozy”□□□13000□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□'un'□'regarded'□□□□□□□□□□'sitdown'□□□□□□□□□□□□□□□□□□□□□□□□157/159=98.7%□□□□□□□□

□□□□□□□□□□“unregarded”□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□1005493□□□□□□7557□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

□□□Pw['unregarded']=7557
□□□segment
□'faroutintheunchartedbackwatersoftheunfashionableendofthewesternspiral
armofthegalaxyliesasmallunregardedyellowsun'□
['far'□'out'□'in'□'the'□'uncharted'□'backwaters'□'of'□'the'□'unfashionable'□'end
'□'of'□'the'□'western'□'spiral'□'arm'□'of'□'the'□'galaxy'□'lies'□'a'□'small'□'unregard
ed'□'yellow'□'sun']

[illegible]

[illegible]“sit”□“down”[illegible]0.003%□0.04%[illegible]“sitdown”[illegible]“sit down”[illegible]100[illegible]

P(W1□n□=Πk=1□nP(Wk|Wk-1□

[illegible]100000[illegible]250000[illegible]Count(sit down)/Count(sit)[illegible]P(down|sit)[illegible]cPw[illegible]

```
def cPw(word,prev)□
"The conditional probability P(word|previous-word)."
try□
return P2w[prev+''+word]/float(Pw[prev]□
except KeyError□
return Pw(word)
P2w=Pdist(datafile□'count2w'□□N□
```

[illegible]cPw[illegible]1[illegible]“[illegible]”(supid backoff)[illegible]“to”□“sitdown”□“sit down”[illegible]

```
□□□cPw□'sit'□'to'□*cPw□'down'□'sit'□/cPw□'sitdown'□'to'□
1698.0002330199263
```

□□□□“sit down”□“sitdown”□□□□1698□□□□“sit down”□□□□□

□□□□□□□“to sit”□□□□□□“to sitdown”□□□□□

□□□

□□□

1.□□□□□n□□□□□□□□□□□□□□□□□□□□□Pwords□□□n+1□□□□□□□

□□□□□□□□□□□□□n□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

2.□□□□□□□□□□□□□□□□□□□□Pwords□□□□□□□□61□“blah”□□□□

□□□□□□□5.2×10^{-321}□□□□□□□□□□□□“blah”□□□□□0.0□□□□□□□□□

□□□□□□4.9×10^{-324}□□□□□□□□□□□□□0.0□□□□□□□□□□□□□□□□□□□□

□□□□□□□□□□□□□□□□□□□□□

□□□□□“segment2”□□□2□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

□□cPw□□□□□□□□Pw□□□□□□□□□□□□□“segment2”□□□□□“□□”□□□□

□□□

□□S□□□

□□□□□□□□□□□□□□□□□□memo)□□□□□□□□□□□□□□□□□1□□□□□□□□

□□□

□□□□□□□□□□□□□□□□□□□□□□□□□□2□[1]□□□□□□□□□□□□□□□□□□□□□

□□□□□□□□□□□□□□□□□□

□□□“segment2”□□□□

```
from math import log10
@memo
```

```
def segment2(txt,prev='□S□'□□
"Return(log P(words)□words)□where words is the best segmentation."
if not text:return 0.0□[]
candidates=[combine(log10(cw(first,prev)□□first,segment2(rm,first)□
for first,rem in splits(text)]
return max(candidates)
def combine(Pfirst,first□(Pem,rem)□□
"Combine first and rem results into one(probability,words)pair."
return Pfirst+Prem□[first]+rem
```

---

segment2□□□O(nL)□□□□□□□□□□□□□□□O(L)□□□□□□□□□□□□□□□□O(nL^2□□□□□□□□□Viterbi□□□memo□□□□□Viterbi□□

segment2□□□□“sit down”□□□“unregarded”□□□□□□□□

□□

[1]□□□□□□□□□□□□

□□

□□□26□□□□□□□□□□□□□□□“a”□□□□□□“b”□□□□□□□□□□□□□□□□□(Pthon□□□□maketrans□translate□□□□□□□□□

```
def encode(msg,key)□
"Encode a message with a substitution cipher."
return msg.translate(string.maketrans(ul(alphabet)□ul(key)□□
def ul(text)□return text.upper□□+text.lower□□
alphabet='abcdefghijklmnopqrstuvwxyz'
```

□□n□□□□□□□□n=1□□□“b”□□“a”□“c”□□“b”……□□“a”□□“z”□□□□□□□□□(Cesar)□□□□□□□□50□□□□□□(sift)□□□□□□□□□□□□□□

```
def shift(msg,n=13□□
"Encode a message with a shift(Caesar)cipher."
return encode(msg,alphabet[n□]+alphabet[□n]□
□□□□□□□□
□□□shift□'Listen,do you want to know a secret?'□
'Yvfgra,qb lbh jnag gb xabj n frperg?'
□□□shift□'HAL 9000 xyz'□1□
'IBM 9000 yza'
```

□□26□□□□□□□□□□□□□□□□□□□

□□□□□□□□□□□□logPwords□□□□Pwords□□□□□□□□□□□□□□

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

```
def logPwords(words)□
"The Naive Bayes probability of a string or sequence of words."
if isinstance(words,str)□words=allwords(words)
return sum(log10(P(w)□for w in words)
def allwords(text)□
"Return a list of alphabetic words in text,lowercase."
return re.findall□'[a-z]+'□text.lower□□□
□□□□□□□□□□□□□□□□□□□□□□□□
def decode_shift(msg)□
"Find the best decoding of a message encoded with a shift cipher."
candidates=[shift(msg,n□for n in range(len(alphabet)□]
return max(candidates,key=logPwords)
□□□□□□□□□□□□□□□□□□□□□□□
□□□decode_shift□'Yvfgra,qb lbh jnag gb xabj n frperg?'□
'Listen,do you want to know a secret?'
```

□□□□□□□□□□□□□□□□□□□□□□□□□□□□26□□□□□□□□□□□□□□□□□□

```
Yvfgra,qb lbh jnag gb xabj n frperg?-84
Zwghsb,rc mci kobh hc ybck o gsqfsh?-83
Axhitc,sd ndj lpci id zcdl p htrgti?-83
Byijud,te oek mqdj je adem q iushuj?-77
Czjkve,uf pfl nrek kf befn r jvtivk?-85
Daklwf,vg qgm osfl lg cfgo s kwujwl?-91
Eblmxg,wh rhn ptgm mh dghp t lxvkxm?-84
Fcmnyh,xi sio quhn ni ehiq u mywlyn?-84
Gdnozi,yj tjp rvio oj fijr v nzxmzo?-86
Heopaj,zk ukq swjp pk gjks w oaynap?-93
Ifpqbk,al vlr txkq ql hklt x pbzobq?-84
Jgqrcl,bm wms uylr rm ilmu y qcapcr?-76
Khrsdm,cn xnt vzms sn jmnv z rdbqds?-92
Listen,do you want to know a secret?-25
MjtufO'ep zpv xbou up lopx b tfdsfu?-89
Nkuvgp,fq aqw ycpv vq mpqy c ugetgv?-87
Olvwhq,gr brx zdqw wr nqrz d vhfuhw?-85
Pmwxir,hs csy aerx xs orsa e wigvix?-77
Qnxyjs,it dtz bfsy yt pstb f xjhwjy?-83
Royzkt,ju eua cgtz zu qtuc g ykixkz?-85
Spzalu,kv fvb dhua av ruvd h zljyla?-85
Tqabmv,lw gwc eivb bw svwe i amkzmb?-84
```

```
Urbcnw,mx hxd fjwc cx twxf j bnlanc?-92
Vscdox,ny iye gkxd dy uxyg k combod?-84
Wtdepy,oz jzf hlye ez vyzh l dpncpe?-91
Xuefqz,pa kag imzf fa wzai m eqodqf?-83
```

[illegible]Pwords[illegible]-25[illegible]10^{-25}[illegible]10^{50}[illegible]

[illegible]“I”[illegible]“a”[illegible]“the”[illegible]'[illegible]“s”[illegible]“t”[illegible]shift2[illegible]

```
def shift2(mg,n=13□□
"Encode with a shift(Caesar)cipher,yielding only letters[a-z]."
return shift(just_letters(msg)□n□
def just_letters(text)□
"Lowercase text and remove all characters except[a-z]."
return re.sub□'[^a-z]'□''□text.lower□□□
```

[illegible]

```
def decode_shift2(mg)□
"Decode a message encoded with a shift cipher,with no spaces."
candidates=[segment2(sift(msg,n□□for n in range(len(alphabet)□]
p,words=max(candidates)
return''.join(words)
```

[illegible]

```
□□□shift2□'Listen,do you want to know a secret?'□
'yvfgraqblbhjnaggbxabjnfrperg'
□□□decode_shift2□'yvfgraqblbhjnaggbxabjnfrperg'□
'listen do you want to know a secret'
□□□decode_shift2(sift2□'Rosebud'□□
'rosebud'
□□□decode_shift2(sift2□"Is it safe?"□□
```

```
'is it safe'
□□□decode_shift2(sift2□"What's the frequency,Kenneth?"□□
'whats the frequency kenneth'
□□□msg='General Kenobi:Years agO'you served my father in the Clone Wars□now he begs you to help him in his struggle against the Empire.'
□□□decode_shift2(sift2(mg)□
'general kenobi years ago you served my father in the clone wars now he begs you to help him in his struggle against the empire'
```

□□26□□4×10^{26}□□□□□□□□□26□□Simon Singh□□□□□□□□The Code Book□□□□□□□□□□□□□□□□□□□□□□

1.□□□□□□□□□□□□□□□□□□□□□□□□“e”□□□□□□□□□□□□□“z”□□

2.□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

3.□□□□□“the”□“and”□“of”□□□□□□□“a”□“I”□

4.□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

5.□□□□□□□□□□□□□□□□□□□□“your faithful servant”□□□□□□

6.□□□□□□□□□□□“abbccddedf”□□□□□“bookkeeper”□□□□□□□□□□□□□□□□

□□□□□□□□□□□□3□□6□□□□□1□□2□□□□26□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□4□□5□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□6□□□□□□□□□□□□□

[illegible]N[illegible]3[illegible]count_3l[illegible]26^3=17576[illegible]10[illegible]

```
the 2.763%fzq 0.0000004%
ing 1.471%jvq 0.0000004%
and 1.462%jnq 0.0000004%
ion 1.343%zqh 0.0000004%
tio 1.101%jqx 0.0000003%
ent 1.074%jwq 0.0000003%
for 0.884%jqy 0.0000003%
ati 0.852%zqy 0.0000003%
ter 0.728%jzq 0.0000002%
ate 0.672%zgq 0.0000002%
```

[illegible]

```
def logP3letters(text)□
"The log-probability of text using a letter 3-gram model."
return sum(log10□P3l(g)□for g in ngrams(text□3□□
P3l=Pdist(datafile□'count_3l'□□
P2l=Pdist(datafile□'count_2l'□□##We'll need it later
```

[illegible]4×10^{26}[illegible]“[illegible]”(hll climbing)[illegible]x[illegible]

□□□□□□□□□□□□□□□□“□□”□□□□□□□□□x□□□

□□□□□□□□□□□□□□□□□□x□□□□□□□□□□□f□□□□□□□□□□□□□□□neighbors□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□x□□□□□□

```
def hillclimb(x,f,neighbors,steps=10000□□
"Search for an x that miximizes f(x)□considering neighbors(x)."
fx=f(x)
neighborhood=iter(neighbors(x)□
for i in range(steps)□
x2=neighborhood.next□□
fx2=f(x2□
if fx2□=fx□
x,fx=x2□fx2
neighborhood=iter(neighbors(x)□
if debugging:print'hillclimb□'□x,int(fx)
return x
debugging=False
```

□□□f=logP3letters□□□□□□□□□□□□□□x□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□segment2□□□□□□□□□□□□□□□□□□□□logP3letters□

```
def decode_subst(msg,steps=4000□restarts=20□□
"Decode a substitution cipher with random restart hillclimbing."
msg=cat(allwords(msg)□
candidates=[hillclimb(encode(msg,key=cat(shuffled(alphabet)□□□
logP3letters,neighboring_msgs,steps)
for_in range(restarts)]
p,words=max(segment2(c)for c in candidates)
return''.join(words)
```

```
def shuffled(seq)□
"Return a randomly shuffled copy of the input sequence."
seq=list(seq)
random.shuffle(seq)
return seq
cat=''.join
```

□□□□□□□□□□□□□□□□neighboring_msgs□□□□□□□□□□□□□□□□□□□□□□□□□“□□”□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□“jq”□□□□0.0001%□□□□□□□□□□□□□□□□“in”□“th”□□□□5□□□□□□□□□□□□□□□□□“jq”□□□□□□“j”□□□□□□□□□□□□□□□□□□□“q”□□20□□□□□□□□□□□□□□“□□”□□□□□□□□□□□□

```
def neighboring_msgs(msg)□
"Generate nearby keys,hopefully better ones."
def swap(a,b□□return msg.translate(string.maketrans(a+b,b+a□□
for bigram in heapq.nsmallest□20□set(ngrams(msg□2□□□P2l□□
b1□b2=bigram
for c in alphabet□
if b1==b2□
ifP2l(c+c□□P2l(bigram)□yieldswap(c,b1□
else□
if P2l(c+b2□□P2l(bigram)□yield swap(c,b1□
if P2l(b1+c□□P2l(bigram)□yield swap(c,b2□
while True□
yield swap(random.choice(alphabet)□random.choice(alphabet)□
```

□□□□□□□□□□□□□□□□□□□□Robert Raynard□□□Secret Code Breaker□(Sith□Daniel□□http□//secretcodebreaker.com)□□□□□□□□□□□□□□□□□□□□□□□□□“□□”□□□□□□

□□□msg='DSDRO XFIJV DIYSB ANQAL TAIMX VBDMB GASSA QRTRT CGGXJ MMTQC IPJSB AQPDR

SDIMS DUAMB CQCMS AQDRS DMRJN SBAGC IYTCY ASBCS MQXKS CICGX RSRCQ ACOGA SJPAS
AQHDI ASBAK GCDIS AWSJN CMDKB AQHAR RCYAE'
□□□decode_subst(msg)
'it is by knowing the frequency which letters usually occur and other distinctive
characteristics of the language that crypt analysts are able to determine the
plain text of a cipher message j'

□□□□□□□□□“crypt analyst”□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□1300□□□□□□□□□□□□□□□□□□□□□□“E”□□□□□□□□□□□□□□□□□□□□□□

□□□□□□□□□□□□□□□□□Baron August Schluga□□□□□□□

□□□msg='NKDIF SERLJ MIBFK FKDLV NQIBR HLCJU KFTFL KSTEN YQNDQ NTTEB TTENM QLJFS
NOSUM MLQTL CTENC QNKRE BTTBR HKLQT ELCBQ QBSFS KLTML SSFAI NLKBR RLUKT LCJUK
FTFLK FKSUC CFRFN KRYXB'
□□□decode_subst(msg)
'english complaining over lack of munitions they regret that the promised support of
the french attack north of arras is not possible on account of munition insufficiency wa

---

□□□□1992□□□□□□□□□□□(KB)□□□□□□□□□(CA)□□Aldrich Ames□□□□□□Aldrich Ames□1994□□□□□□□

□□□msg='CNLGV QVELH WTTAI LEHOT WEQVP CEBTQ FJNPP EDMFM LFCYF SQFSP NDHQF OEUTN
PPTPP CTDQN IFSQD TWHTN HHLFJ OLFSD HQFED HEGNQ TWVNQ HTNHH LFJWE BBITS PTHDT
XQQFO EUTYF SLFJE DEFDN IFSQG NLNGN PCTTQ EDOED FGQFI TLXNI'
□□□decode_subst(msg)
'march third week bridge with smile to pass info from you to us and to give
assessment about new dead drop ground to indicate what dead drop will be used next
to give your opinion about caracas meeting in october xab'

□□□□1943□□□□U-Boat□□□□□□□□□□□□□□□□□□□□□□□□□□□□

```
msg='WLJIU JYBRK PWFPF IJQSK PWRSS WEPTM MJRBS BJIRA BASPP IHBGP
RWMWQ SOPSV PPIMJ
    BISUF WIFOT HWBIS WBIQW FBJRB GPILP PXLPM SAJQQ PMJQS RJASW LSBLW
GBHMJ
    QSWIL PXWOL'
□□□decode_subst(msg)
'a cony ov is headed northeast take up positions fifteen miles apart between
point
    yd and bu maintain radio silence except for reports of tactical importance x
abc'
```

□□□□□□“y”□“x”□□□□□□“conyov”□□□“convoy”□□□“point yd”□□“point vd”□□□segment2□□□□□□□□□□□□□□□□□

□□□□

□□□□□□□□□□□□□□□□□□□w□□□□c□□□□□□□□□□□□□□□□□w□“acomodation”□c□□□“accommodation”□□□w□“the”□□□c□□□“the”□□

□□□□□□□□□□□c□□□□□□P(c|w□□□□□□□□□□□□□□□□□w=“thew”□□c□□□□□□“the”——□□□□□□□□□□□□□□□□□□□“e”□□□□“w”□□□□□□“thaw”——□□□□□□□□□□□□□“the”□3□□□□□□□□□□□□□□□□□□□□□□“thew”□□□□□□□□□□□□□□□“threw”□□□□“Thwe”□□□□□□□□□□□□□□□□□□□c□□□□□□□w□c□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□(a hoc)□□□□□□□□□□□□□□□□□□——□□□□——□□□□□□□□□□□□□□□□□□□□

argmaxc P(c|w□=argmaxc P(w|c□P(c)

□□P(c)□□c□□□□□□□□□□□□□□□P(w|c□□□□□□□c□□w□□□□□□□□□□□□□□□□□□□□□□□□□c□□□□□□□□□□□□□□□w□□□□□□□□□□□□□□□□□□□□□□□□□——□□□□□□□□□□□□□□□□□□□□□

□□□□□□□□□□□□□□□□□□□□□□Roger Mitton□http□//www.dcs.bbk.ac.uk/□ROGER/corpora.html□□□□□1□□□c,w□□□□□□□□□□□□□□□□□□□□P(w=thew|c=thaw)□□□4

□□

□□“th”□“w”□□□□□□□□P(w=e|c=a□□□□□□□□□“a”□□□“e”□□□□□□

□□□“consistent/consistant”□“inseparable/inseperable”□□□□

□□□“a”□□□□“e”□□□□□□□□□□□□□

□□□□□□□□□□□w=thew□□c□□□□□□□□□□□□□thew□□□□□□□□□□□

□□□□□□□□□□□□□□□1□□□□□“thew”□□“w”□□□“the”□2□□□“r”□□

□“threw”□□□□□□□□□□□□□□□□□□□□□□□3□□“a”□□“e”□□□□□□□4□□□□

□□□□□□□□□□□“ew”□□□“we”□□□□□□□□□□□□□□□□□1□□□□□□□□□□□□□□

□□□2□□□□□□□w□c□□□w|c□□□P(w|c□□□□P(c)□□□□□□□□□□□□□□□□□

□□□□□□□□

w	c	w | c	P(w | c)	P(c)	10^9 P(w | c) P(c)
thew	the	ew | e	0.000007	0.02 144.	144.
thew	thew		0.95	0.00000009	90.
thew	thaw	e | a	0.001	0.0000007	0.7
thew	threw	h | hr	0.000008	0.000004	0.03
thew	thwe	ew | we	0.000003	0.00000004	0.0001

□□□□□□□□□“the”□□□□□□□□□□□P(c)□□□Pw□□□□□□□P(w|c□□□□□

□□□□□□□□□□□Pedit□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

□□Pedit□'ew|e'□=0.000007□□□□□□□□□□□□□□□□□□□□□□□□

□“hallow”□□□“hello”□□□□a|e□□□□□ow|o□□□□□□□□□□□□□□□□□□□□

a|e+ow|o□□□ow|o+a|e□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

□□□□□□□□□□

□□□□□□□□□Pedit□''□□□□□□□□□□□□□□□□□□c□□□□□□c□□
□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□20□□□□□□□
□□□□□□□□□□□□□□□50□□□□□□□□□□□□□w=“thew”□P(w|c□□
□□□0.98.□□□□0.95□□□“thew”□□□□□□□□

□□□□□□□□□□□□□□□□□□□□□correct□corrections,correct□
□□□□□□□□□□□□corrections□□□correct□□□□□□□□□□□□□□□□□
□□□□□□□□□□□□□□□□□□□□□□□□P(w|c□P(c)□□□□□□

```
def corrections(text)□
"Spell-correct all words in text."
return re.sub□'[a-zA-Z]+'□lambda m:correct(m.group□0□□□text)
def correct(w)□
"Return the word that is the most likely spell correction of
w."candidates=edits(w).items□□
c,edit=max(candidates,key=lambda(c,e□□Pedit(e)*Pw(c)□
return c
P(w|c□□□□□Pedit□□□□□
def Pedit(edit)□
"The probability of an edit□can be''or'a|b'or'a|b+c|d'."
if edit==''□return□1.-p_spell_error)
return p_spell_error*product(P1edit(e)for e in edit.split□'+'□□
p_spell_error=1./20.
P1edit=Pdist(datafile□'count1edit'□□##Probabilities of single edits
```

□□□□□□□□□□□□□□□□□□□□□{□□□□}□□□□dict□□□□□□□□□
□□□□□□□□□□□□□□□□□□□□□□□□□“e”□□“l”□□□“l”□□□□□
□“tel”□□“tell”□□□□□□□□□□□□□edits□□□□□□□□□□□□□□□□□□
□□□□□□□□□□□□4□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□
□□□□□□□□□□□□□http□//norvig.com/spell-correct.html□□□□
□□□□□□□□□□□□□“accommodations”□□□□233166□□□□□□□□

11□□□

□□□edits□□□□□□□□□□□□□□□□□□□hd□tl)□□□□□□□□□□□□□□□□□□□□□□□□□□□

□□□□dict□□□□□□

```
def edits(word,d=2□□
"Return a dict of{correct:edit}pairs within d edits of word."
results={}
def editsR(hd,tl,d,edits)□
def ed(L,R□□return edits+[R+'|'+L]
C=hd+tl
if C in Pw□
e='+'.join(edits)
if C not in results:results[C]=e
else:results[C]=max(results[C]□e,key=Pedit)
if d□=0□return
extensions=[hd+c for c in alphabet if hd+c in PREFIXES]
p=(h[-1]if hd else'□'□##previous character
##Insertion
for h in extensions□
editsR(h,tl,d-1□ed(p+h[-1]□p□□
if not tl:return
##Deletion
editsR(hd,tl[1□]□d-1□ed(p,p+tl[0]□□
for h in extensions□
if h[-1]==tl[0]□##Match
editsR(h,tl[1□]□d,edits)
else□##Replacement
editsR(h,tl[1□]□d-1□ed(h[-1]□tl[0]□□
##Transpose
if len(tl)□=2 and tl[0]□=tl[1]and hd+tl[1]in PREFIXES□
editsR(hd+tl[1]□tl[0]+tl[2□]□d-1□
ed(tl[1]+tl[0]□tl[0□2]□□
##Body of edits□
editsR□''□word,d□[]□
return results
PREFIXES=set(w[□i]for w in Pw for i in range(len(w)+1□□
```

□□□□□edits□□□□□□□□

```
□□□edits□'adiabatic'□2□
{'adiabatic'□''□'diabetic'□'□a|□+a|e'□'diabatic'□'□a|□'}
```

□□□□□□□□□□□□□□

```
□□□correct□'vokabulary'□
'vocabulary'
□□□correct□'embracable'□
'embraceable'
□□□corrections□'thiss is a teyst of acommodations for korrections
of mispellings of particuler wurds.'□
'this is a test of acommodations for corrections of mispellings
of particular words.'
```

15□□□□□13□□□□□□□□□□“acommodations”□“mispellings”□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□“mispellings”□□□□□□□□18543□□□□□□□□“misspellings”□□□□50□□□□□□□□□□□□□□□□□□□□□□□□□“thew”□□□96759□□□□□□□□□□□□□□□□□□□□

□□□“they're”□□□□□□□□□□□□□“in they're words”□□□□□□□□“their”□□□□□□□□□□□□□□□□□□□□□

□□□□□□□□□□□□□□□□□□□□□□□□□□“misspellings”□

```
misspellings 432354
mispellings 18543
misspelings 10148
mispelings 3937
```

□□□□□□□misspellings□□□□□□□□□□□□□7%□□

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□“□□□”(crcular reasoning)□□“mispellings”□“misspellings”□□□□□□□□

```
mispellings allowed 99
misspellings allowed 2410
mispellings as 50
misspellings as 749
mispellings for 122
misspellings for 11600
mispellings of 7360
misspellings of 16943
```

□□□□□□□□□□□□□□□□□□□□□□□□“misspellings”□□□□□□□□□□□□“mispellings”□□□□□□□□□□□□□□——□□□□□□□□□CPU□□□□□□□□□□□□□□□□□□□□□□□□□□□

□□ht□//Dig□□□□□□□Dig□□□□□□□□□□□□□□□□□□ht□//Dig□□metaphone□□□□□□□□□□□□□□□□□□“tough”□“tuff”□□□□□“TF”□□□□□□□□□□□□□□□□□“G”□metaphone□□□□

```
case'G'□
/*
*F if in-GH and not B--GH,D--GH□
*-H--GH□-H---GH else dropped if
*-GNED□-GN□-DGE-□-DGI-□-DGY-
*else J if in-GE-□-GI-□-GY-and
*not GG else K
```

```
*/
if□□*□n+1□□='G'||vowel□*□n+2□□□□□
□*□n+1□□='N'||□*□n+1□□□
□*□n+2□□='E'||
*□n+3□□='D'□□□□□
□*□n-1□□='D'||□frontv□*□n+1□□□□
if(frontv□*□n+1□□□□*□n+2□□='G'□
key□□'J'□
else
key□□'K'□
else if□*□n+1□=='H'□□□noghf□*□n-3□□□□
*□n-4□□='H'□
key□□'F'□
break□
```

□□□□□□□□“TOUGH”□□“GH”□□□“F”□□□□□□□“BOUGH”□□“GH”□□□“F”□□□□□□□□□□□□□□□□□□□□□“OUGHT”□□□□“COUGH”□“HICCOUGH”□□□“iPhone”□□□——□□correct□□□50□□□□□□□ht□//Dig□□□□□□□1500□□□□□□□Bill Gate□□□“□□□□□□□□□□□□□□□□□□□□□□□□□□□□□”Gates□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

□□□□□□□□□□□□□□□□□□□□□□□□□□□□(Ltvian)□□□□□□□□□□metaphone□□

□□□□

□□□□□□□□□□□□□□□□□□□□□□

□□□□

□□Web□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□HTML□□□□□

HTTP□□99%□

□□□□□□□□□□□□□□

□□□□□□□□□□□□□□□1000□□□N□□□□□□□□“10000000.00 will be released”□“our country Nigeria”□□□□□□□□□□□□□□□N□□□□□□“v1agra”□□□99%□□□□□□□□□□□□□□□□□/□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□/□□□□□□□□□□□□□□□□□□□□□□□□/□□/□□/□□□□□□□□□□□□□□“□□/□□/□□□”□

□□□□□□□□□Stylometry)

□□□□□□□□□□□□□Federalist Papers□□□2004□□□□□□□□□□□□□□71%□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

□□□□□□□DNA□□□

□Vernor Vinge□□□□□□Rainbows End□(Tr Books□□□□ Librareome□□□E-Puzzler□□□□□□□□4500□□□□□□□□□□□□□□□□□□□Stasis□□□shotgun□□□□□□□□□□DNA□□□□□□□□□“□□□□□□”□□□□□□□Web□□□□□□□□□□□□□□□□

□□□□

G□□□N□□□□□□□□□□□□□□□□□□□□□(f)□□□□□(e)□□□□□□□□□□□□□□□□□□□□□□□□□□

best=argmaxe P(e|f□=argmaxe P(f|e□P(e)

□□□P(e)□□□□□□□□□□□N□□□□□□□□□□□□P(f|e□□□n□□□□□□□□

[illegible]

[illegible]

[illegible]——[illegible]“choosespain.com”[illegible]N[illegible]

[illegible]“[illegible]”[illegible]

[illegible]

[illegible](Ehaustive)

[illegible]26[illegible]

[illegible](Garanteed)

[illegible]2^n[illegible]

[illegible](Huristic)

[illegible]

[illegible]“[illegible]”(i the wild)[illegible]N[illegible]Web[illegible]1[illegible]4[illegible]——[illegible]——[illegible]

[illegible]——[illegible]——[illegible]——[illegible]1[illegible]2[illegible]3[illegible]——[illegible](a hoc rule)[illegible]

[illegible]

[illegible]

[illegible]Darius Bacon[illegible]Thorsten Brants[illegible]Andy Golding[illegible]Mark Paskin[illegible]Franco Salvetti[illegible]Casey Whitelaw[illegible]

[illegible]15[illegible] [illegible]DNA[illegible]

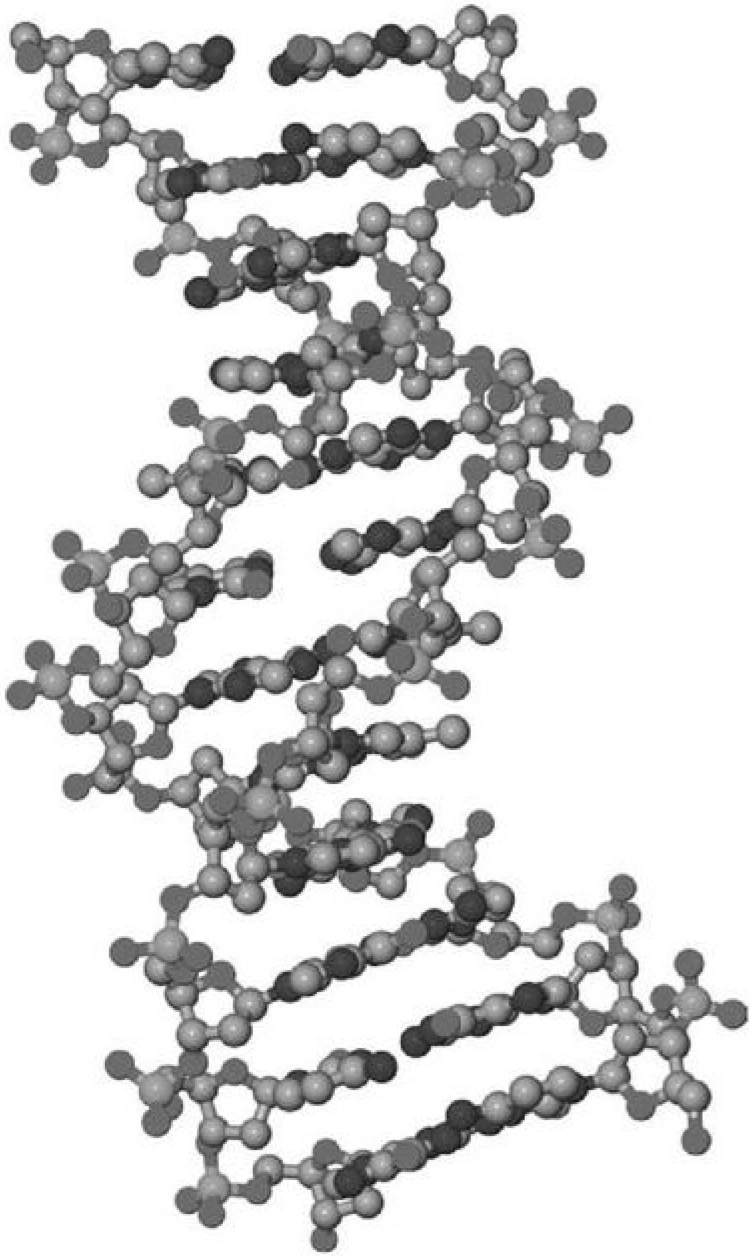
MattWood[illegible]Ben Blackburne

DNA[illegible]20[illegible]DNA[illegible]DNA[illegible]DNA[illegible]DNA[illegible]

□DNA□□□□

□□□□□□□□□□DNA□□□DNA□□□□□□□□□

□□X□□□□□□□X□□□Y□□□□□□□□□□□□□□□DNA□□□□□□□□□□□□□□□□□□15-1□□□



□　15-1□DNA□□□□□□□□□POV-Ray□□□PDB□□1BNA□□□doi□10.2210/pdb1bna/pdb□□□□51□

DNA□□□□□□□□□□□□□□DNA□□□□□□□——□□□□□□□□□□□□□□□□□□□□□□□A□G□C□T□□□□□□□□□□□□□□□□□□□□□□□□□□DNA□□3G□3×10^9□□□□□□□□□□□□□□□□□□□

DNA□RNA,RNA□□□□

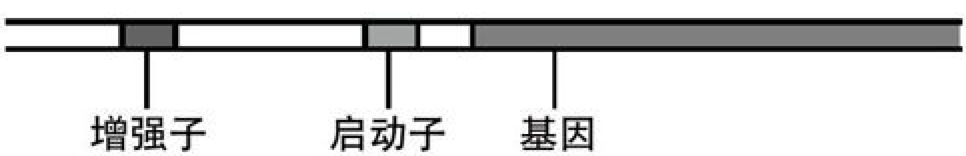
DNA□□2%□□□□□□□□□□□□□□□□□□□□DNA□□□

□□□□□□□□□□□□□□□DNA□□□RNA□DNA□□RNA□□□□□□□□□□□□RNA□□□□□□□□□□□□□□□□□□□□□RNA□□□□□□□□□□□□□□□□□

□□□□□DNA□□DNA□□□□“□□□□□”□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

□□□

□□□□□□□□□□□□□□□□□□□□□□□□□□□15-2□□□



□ 15-2□□□□□□DNA□□□

基因的多态性进行了52。

　　另类的：Hack你的DNA密码

　　近日，知名出版公司O'Reilly推出“Hacks”系列的最新作品——《DNA Hacks》——就收录了100条读者在家中也能自行研究基因的小技巧，其中一条Hack就是关于如何检测自己对一夫一妻制忠诚度的。文中介绍，人类体内存在一种名叫G蛋白偶联受体（G-protein coupled receptors,GPCR)的膜受体。它广泛存在于人类的大脑中，GPCR的种类有许多。其中一种GPCR受体的基因拥有两种不同的版本：“长的”和“短的”。对DNA做了分析后，研究者发现，拥有长版本GPCR受体基因的人会对伴侣更加忠诚。作者通过Email接受采访时称hack也有另外一面，其实拥有长版本的“V1Agr4”基因的人也许不是很忠诚。不过，hack以后，结果也许可以证实“你是个好男人”，或者给你一个明智的提示。

　　书中的内容是对DNA元数据(mtadata)进行深入研究的结晶，让读者可以用最简单的方法了解自己的DNA，无需专业知识，无需昂贵仪器，你需要的只是一些简单的化学试剂和日常用品。书中的方法都是经过作者亲身验证的，所有的实验都可以在家中完成。而且书中提供了详细的操作步骤，读者可以按照步骤一步一步地完成实验。除此之外，书中还介绍了很多有趣的实验，例如检测自己是否是一个天生的运动员、自己的祖先来自哪里等等。当然，这些实验的结果并不是绝对准确的，读者需要理性看待——这本书的作者也提醒读者，《DNA Hacks》中的实验结果仅供参考，不能作为医学诊断的依据。如果读者对自己的基因有疑问，最好还是去专业的医疗机构进行检测。总之，这是一本非常有趣的书，值得每一个对基因感兴趣的人阅读。

　　结论

　　综上所述，基因在人类的婚姻关系中起着重要的作用。通过对基因的研究，我们可以更好地了解人类的行为，从而更好地处理人际关系。当然，基因并不是决定一切的因素，环境和教育也起着重要的作用。

□□
□DNA□□□□□

□□

□□SMART□□□□□□□□□□□□□□□□□□□□RAID1□□□DNA□□□□□□□□□□□□□□□DNA□□RAID1□□□□□□□□□□□□□□□□□□□□□□□□

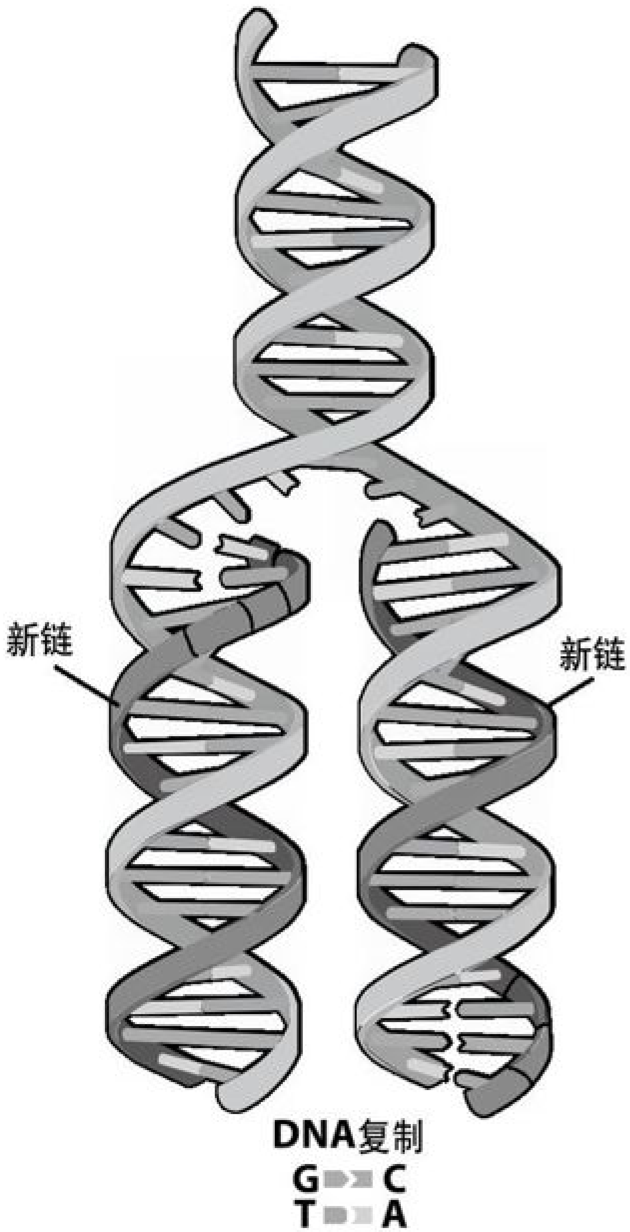
□□□DNA□□□□□□□□□□□□□□□□□□□SMART□□□□□□□□□□□□□□□5000□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□100□□□□□□□□□□□□□□□□□□□□□□□□□□□

□□

□□□□□□□□DNA□□□□□□□□□□□□□□□□□□□□□□□DNA□□□□□□□□□□□□□□□A□G□□□C□T□□□□□□□□□□□□□□□□□□□□□□□□□□[1]□□□□□DNA□□□□□□□□□□□□□□□“□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□”

[illegible]DNA[illegible]DNA[illegible]

[illegible]15-3[illegible]



[illegible] 15-3 [illegible]DNA[illegible]DNA[illegible]

[illegible]http[illegible]//genome.gov/glossary.cfm)

[illegible]

□□□□□□□□□□□21□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□
□□□□□□□□□□□□□□□□□□7bit□□□□□□□□□□7bit ASCII□□□□□□□□□□
□□

1961□□□□□□□□□[2]□□□□□□□□□□□□□□□□□ASCII□□□□□□□90□□
□□□□□□□□□□□7□□□□□2^7=128□□□□□□□□□□□□□□□□□DNA□□□□□□
□□□□□□□□21□□□□20□□□□□1□□□□□□□□□□□□□□□□3□□□□4^3=64□
□□□15-1□□

		T	C	A	G		
	T	Phe	Ser	Tyr	Cys	T	第一个碱基
						C	
		Leu		终止密码子	终止密码子	A	
					终止密码子	G	
	C	Leu	Pro	His	Arg	T	
						C	
				Gln		A	
						G	
	A	Ile	Thr	Asn	Ser	T	
						C	
				Lys	Arg	A	
		Met				G	
第一个碱基	G	Val	Ala	Asp	Gly	T	
						C	
				Glu		A	
						G	

□□□□□□□□□ACT□□□□□□□□□□□□□□□□□□ACT□□□□□□□□□□□□□□
□□□□□□□□□□□□□□□□□□□□□□□□□□□□64□□□□□20□□□□□□□□□□□□□
□□□□□□□□□□□□□□□□□□□——□□□□□□□□□□□□□□□□□□

A T

"DNA" DNA

Alu 300 RNA DNA (Sanley Kubrick) The Shining "" [3] Alu 10%

DNA

"" "" [4]

DNA A T DNA

DNA DNA DNA DNA

[illegible]——[illegible]

[illegible](nll model)[illegible]“[illegible]”[illegible]

[illegible][5][illegible](Gaph)[illegible]

[1][illegible]·[illegible]·[illegible](Jmes D.Watson)[illegible]·[illegible](Fancis Crick)[illegible]DNA[illegible]1962[illegible]

[2][illegible]·[illegible](Sdney Brenner)[illegible]2002[illegible]

[3]“All work and no play makes Jack a dull boy.”[illegible]The Shining[illegible]

[4][illegible]·[illegible](Gegor Mendel)[illegible]

[5]Reference Human Genome[illegible]

DNA□□□□□

□□□□□□□□DNA□□□□□□□□

char□3*10^6□human_genome□

□□□□□□□□□□30□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□30□□

□□□13□□□2001□2□□□10Gb□□□□□□□□□iPod□□U□□10G□□□□□□□□□□□□□□□□□

□□□□□□□□2008□□□□□□□□□□□□□□□□□(htp□//ensembl.org)□□□[1]□□□□□□□□□□□Sanger□□□□IT□□□□□□□□□□□□□□□□□□□DNA□□□□□□□□

□□□30□50□□□□□□□□□□□□□□□30□50□□□□□□□□□□□□□□□□□□□□□□□□□□□□30□□Kb□□□□□□□□□□□□□□□□□□□□□□□

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□DNA□□□□
□□□

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□
□□□□□□□□□□□□□□□□□Sanger□□□□□□50Tb□□□□□□□□
50Tb□

□□□□□□

□□□□□□□□□□□□□□□□□□□100+□□□□□□□□□□□□□□□
□□□□□□□□□□□□□□□□□□□□□□□□□

□□□□□□□□□□Sanger□□□□□□□□□□□□□□□□□□□□
Sanger□□15□□□□□□□□□□□□□□□17000□□□□□□PB□□□
□□□□12□18□□□□□□□□□□□□□15□□□□□

□□□□□□□□□□□□□□□□□□□□6□□□□□□□□□□□□□□□□□
□□□□□□□□□□□□□□□□□□□□□□□□□

□□□

□□□□□□□□□□□□□□□□□□□□□□□□□1Mb□□□□□□□□□
□□□□□□□□□□□□□□

□□□□□□□□□□□□□□□□□□□□DNA□□□□□□□□□□□□□□□
□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

□□□Kb□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□
□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□
□□□□□□□□□□□□□□□□□□□□

(sngle nucleotide polymorphisms,SNP) (cllections of repeating units,CNV) 30

DNA

DNA

DNA

A T,C G

DNA 15-4

148

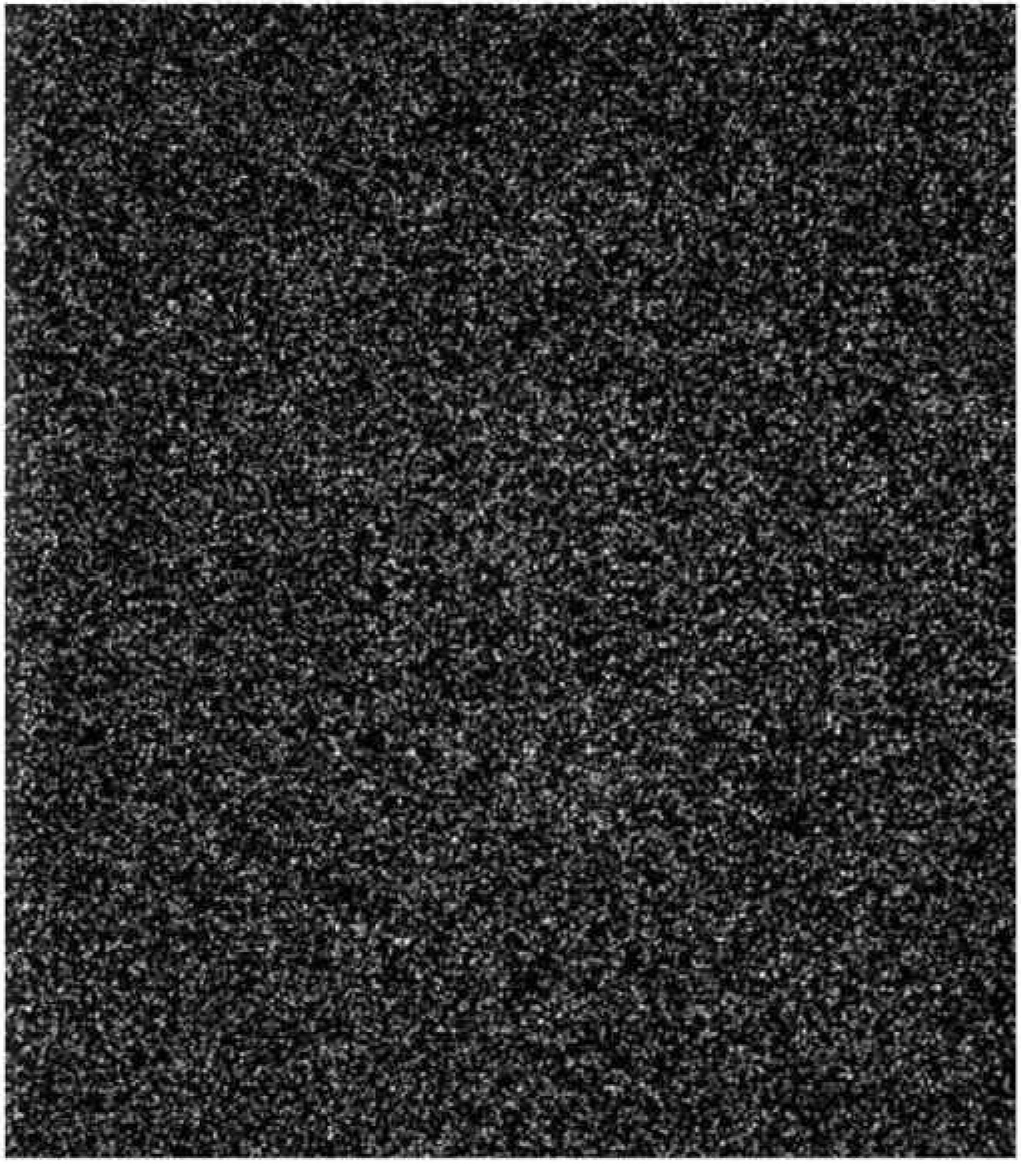
DNA DNA 330 48840

8 390720 DNA

781440

1Tb

fastq SRF



□ 15-4□Illumina[2]GA2□□□□□□□□□□□□□□DNA□□□□□□□□

□□□□

□□□□□□□□□□□□□□□□□□□□□□□□□□Sanger□□□□□□□□□

□□□□□□□36□□□□□□100□□□□□□□□□□Sanger□□□□□□□□□□

□□□□□□□□□□□□□□□2334320□□□□□□□40□□□□□□□□□□□□□

[illegible]9300[illegible]

[illegible]75Tb[illegible]

[illegible]

[1]Wellcome Trust Genome Campus[illegible]Hinxton[illegible]

Sanger[illegible]

[2]Illumina[illegible]DNA[illegible]

[illegible]

[illegible][1][illegible][2][illegible]20[illegible]Sanger[illegible]35[illegible]Illumina GA2[illegible]10[illegible]Hinxton[illegible]

Sanger[illegible]

[illegible]Sanger[illegible]Sanger[illegible]Sanger[illegible]

[illegible]

[illegible]Sequencescape[illegible]Sequencescape[illegible]Rails[illegible]MySOL[illegible]
(htp[illegible]//www.sanger.ac.uk/[illegible]

[illegible]1G[illegible]Sequencescape[illegible]

[illegible]DNA[illegible]

[illegible]DNA[illegible]“[illegible]”[illegible]

[illegible]DNA[illegible]

[illegible]

[illegible]DNA[illegible]

[illegible]

DNA[illegible]DNA[illegible]

[illegible]

[illegible]

[illegible](dta schema)[illegible]Sanger[illegible]

[illegible]

[illegible](Dscriptors)[illegible]

(Fmily)[illegible]UI[illegible]

[illegible]

□□□□□□□□□UI□□

□□□Web□□□

□□Sequencescape□□□□□□□□□□□□□□□□□□□□□□□

http□//www.sanger.ac.uk□□□

□□□□□□□

□Sanger□□□□□□□□□□□□□□□□□□□□□□□□□Web□□

□□□□□□□□□□□□□PC□□□PC□□□□□□□□□□□□□□□□□□□□□□□□□

□□□□□□□□□10G b□□□□□□□□□P C□□□□□□□□□□□□□□Lustre[3]□□□400Tb□□EVA□□□□□□□□□□□□1000□□

[illegible]4[illegible]

[illegible]30Gb[illegible]3[illegible]JPEG[illegible]DNA[illegible]

[illegible]Oracle[illegible]Fuse[4][illegible]Sangery[illegible]

[illegible](Q)[illegible]Sequencescape[illegible]Fuse[illegible]Sequencescape[illegible]

[illegible]Sanger[illegible](htp[illegible]//www.ensembl.org/index.html)[illegible]
(htp[illegible]//trace.ensembl.org/[illegible]
http[illegible]//www.ebi.ac.uk)[illegible]

[1]□1000 Genomes Project□□□□□□□□□□□□□1000□□□□□DNA□□□□□http□//1000genomes.org□

[2]□International Cancer Genome Consortium□□□http□//1000genomes.org□

[3]Lustre□□□□□□□□□□□□□http□//lustre.org/□

[4]http□//fuse.sourceforge.net/

DNA□□□

□□DNA□□□□□□□□□□□□

□□□□□□□□

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□——□□□□□□□□□□□□□□□□□□Enesembl□UCSC)□□□□□□□□□□□□(HpMap,SNPedia)□□□□□23andMe□Navigenics□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

□□□□

□□Pacific Biometrics□Oxford Nanopore□□□□□□□□□□□□□□□□□□□□(mgabase)□□□□□□□□□□(ggabase)□□□□□□□□□□□□□□□□□□□□□□□□□□$5000□□□□$1000□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

□□□□□

□□

[illegible]

[illegible]DNA[illegible]

□□

□□□□□□Wellcome Trust□□□□□□□□□□□□□□□□□□□□□□□

□16□ □□□□□□□□□□

Jean-Claude Bradley□Rajarshi Guha□Andrew Lang□Pierre Lindenbaum□Cameron Neylon□AntonyWilliams□EgonWillighagen

[illegible]

[illegible]“[illegible]”[illegible]“[illegible]”[illegible]

[illegible]“[illegible]”[illegible]“[illegible]”[illegible]“[illegible]”[illegible]“[illegible]”[illegible]

[illegible]KB[illegible]GB[illegible]

[illegible]“[illegible]”[illegible][1][illegible]

[1][illegible]crowdsource[illegible]“[illegible]”[illegible]

[illegible]

[illegible]Jean-Claude Bradley[illegible](Badley 2007[illegible] [illegible]Bradley[illegible] [illegible](Badley[illegible]2008[illegible]Ugi[illegible] [illegible] (htp[illegible]//onschallenge.wikispaces.com/[illegible]

[illegible]“[illegible]”[illegible] (htp[illegible]//en.wikipedia.org/wiki/Open_Notebook_Science)[illegible] Wikispaces[illegible]wiki(http[illegible]//onschallenge.wikispaces.com)[illegible]Gdoc[illegible]Flickr(http[illegible]//flickr.com)[illegible]16-1[illegible]

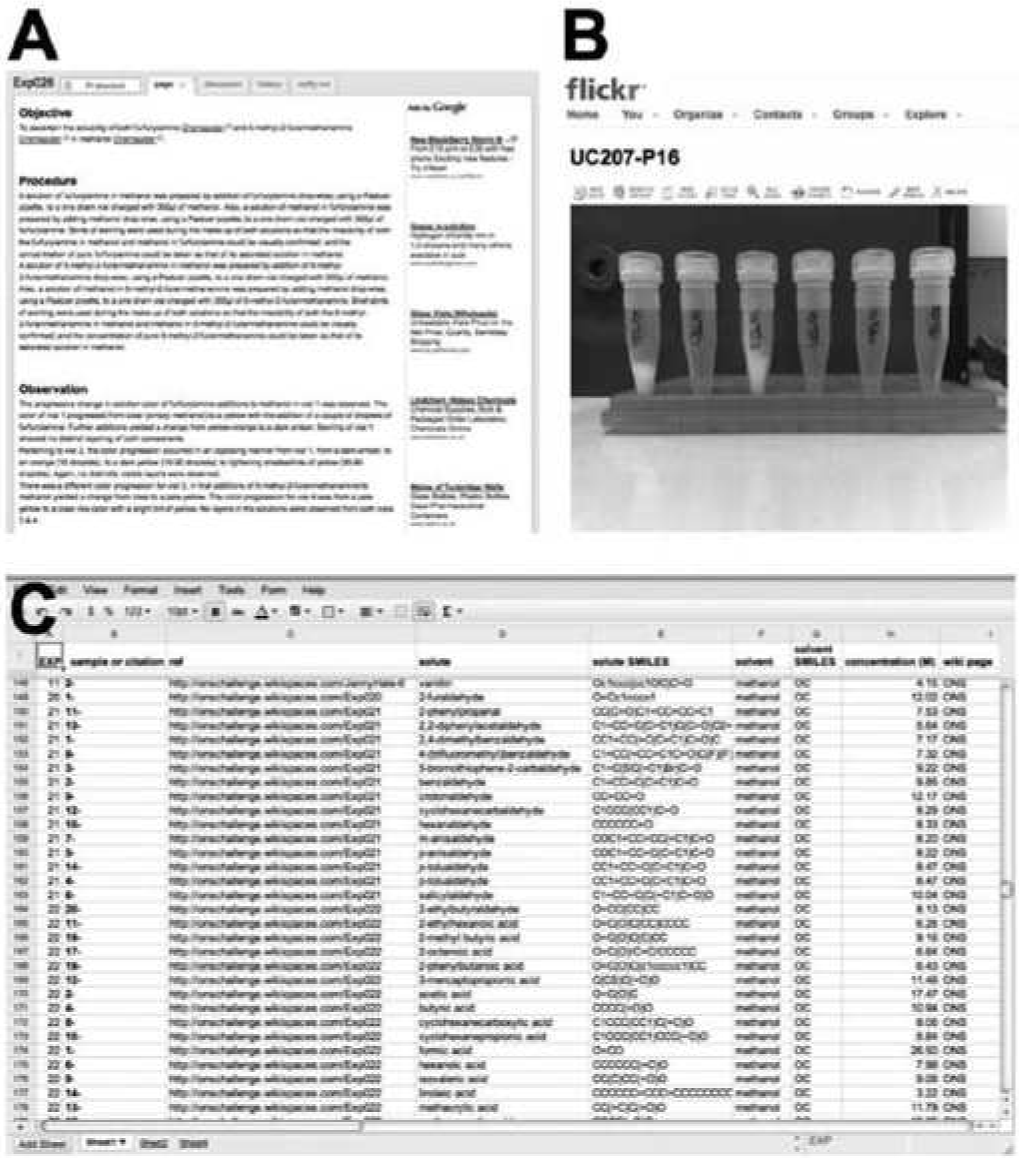


图 16-1[illegible]A[illegible]B[illegible]Flickr[illegible]C[illegible]Gdoc[illegible]53[illegible]

[illegible]Gdoc[illegible]

[illegible]
Gdoc[illegible]
[illegible]
[illegible]API[illegible]
[illegible]

[illegible] 16-2 [illegible]
http [illegible] //oru.edu/cccda/sl/solubility/ugidata.php?solute=4-nitrobenzaldehyde [illegible] solvent=methanol) [illegible]

[illegible]
(htp [illegible] //usefulchem.blogspot.com/2008/12/mechanical-turk-does-solubility-on.html) [illegible]

[illegible] “[illegible]” [illegible]

[illegible] “[illegible]” [illegible]

(htp[illegible]//usefulchem.blogspot.com/2008/11/what-is-solubility-of-vanillin-in.html)[illegible]

[illegible]

[illegible]Web[illegible]

[illegible]

[illegible]

[illegible]

[illegible]CAS[illegible][1]

(htp[illegible]//en.wikipedia.org/wiki/CAS_registry_number)[illegible]

[illegible]CAS[illegible]

[illegible][2][illegible]

IUPAC[3][illegible](IChI[illegible]“INchee”[illegible]

[illegible](htp[illegible]//en.wikipedia.org/wiki/Inchi)[illegible]

[illegible]http[illegible]//www.qsarworld.com/INCHI1.php[illegible]

InChI[illegible]InChI[illegible]——[illegible]

[illegible]InChI[illegible]——[illegible]InChI(Standard InChI)[illegible]

[illegible]InChIKey(InChI[illegible]hash[illegible]InChIKey[illegible]

[illegible]

SMILES(http[illegible]//en. wikipedia.org/wiki/SMILES)[illegible]

[illegible]SMILES[illegible]

[illegible]SMILES[illegible]SMILES[illegible]

□□□□“□□□□□□□□□□□□□□□□□”□□□□□SMILES□□InChI□□
□□□□□□□□□□□□□InChI□□□□□□□

□□□□□□□□□□□□□□□□□□

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□
□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□
□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□
□□

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□
□□SMILES□□□□□GDoc□□□□□□□□□□□□□□□□□□□□□□□
□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□
□□□□□“□□□□□□□”□□□□□□□□□□□□□□□□□□□□□□□□□
□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□
□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□CSV□□□□
□□□□□□□□□□□□□□□□□□□□G□□Doc API□□□□Web□□□
□□□□□□□□“□□”□□□□□

□□□□□□□□□□□□□Web□□□□□□□□□□□□□□□□□□□□
□16-2□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□HTML
□JavaScript□□□□□□□□□□□□□□□□G□□□□API□GDoc□□□□
□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□
□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□G□□

Visualization API□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□
□□
□□□

　　□□Indiana University□□REST□□□□□□□□□□□□URL□□□SMILES□□□□□□□□□□□
□□
□□□HTML□□□□□□□□Web□□□□□□□

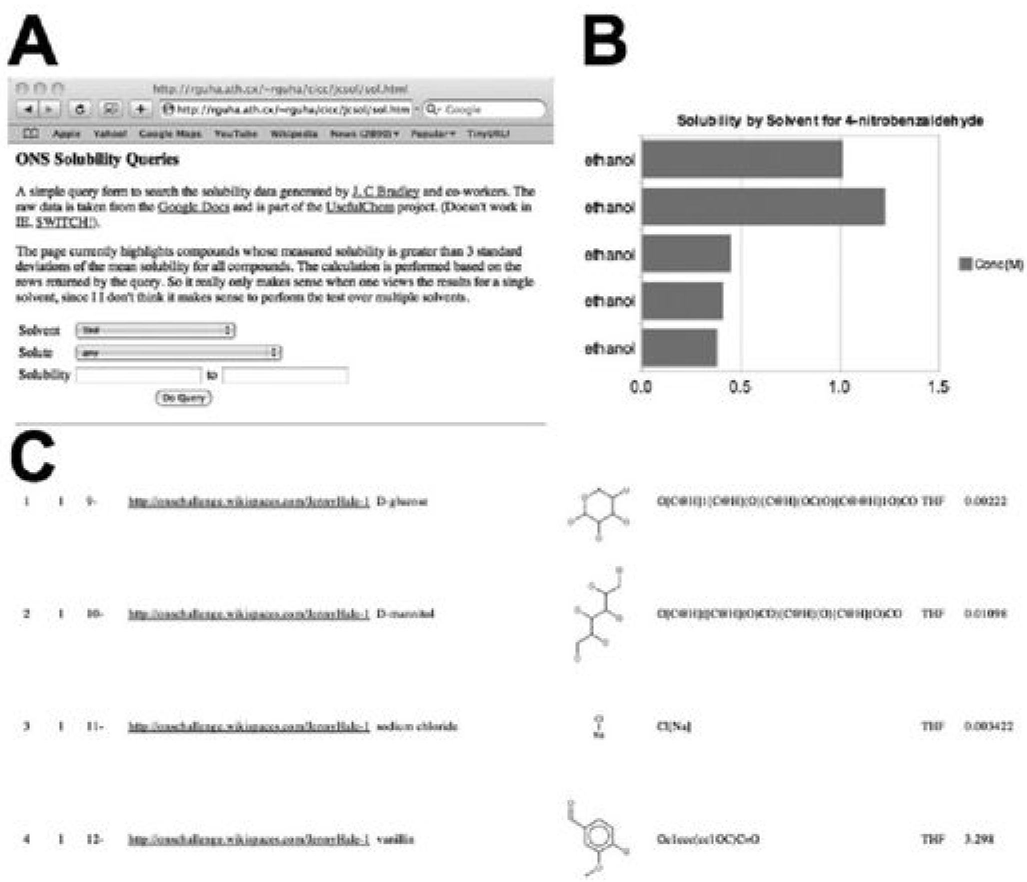


　　图　16-2□□□□□□□□□□□□□□□□□A□□□JavaScript□G□□Doc API□□□□□□□□□□□□□□B□□□□□□□□□□□□□C□□□□□□□□□□□□□□□□□□□□□□
□http□//toposome.chemistry.drexel.edu/□rguha/jcsol/sol.html

[illegible]

54[illegible]

[illegible]

[illegible]

——[illegible]——[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

(Cemical Abstract Service,CAS)[illegible]CAS[illegible]

(htp[illegible]//www.cas.org/newsevents/connections/derivative.html

)[illegible]

ChemSpider(http[illegible]//www.chemspider.com/blog/caschemspid

er-connectivies-and-unintended-collaboration.html)[illegible]

ChemSpider[illegible]“[illegible]”[illegible]Web[illegible]

[illegible]150[illegible]ChemSpider[illegible]

[illegible]Internet[illegible]

[illegible]

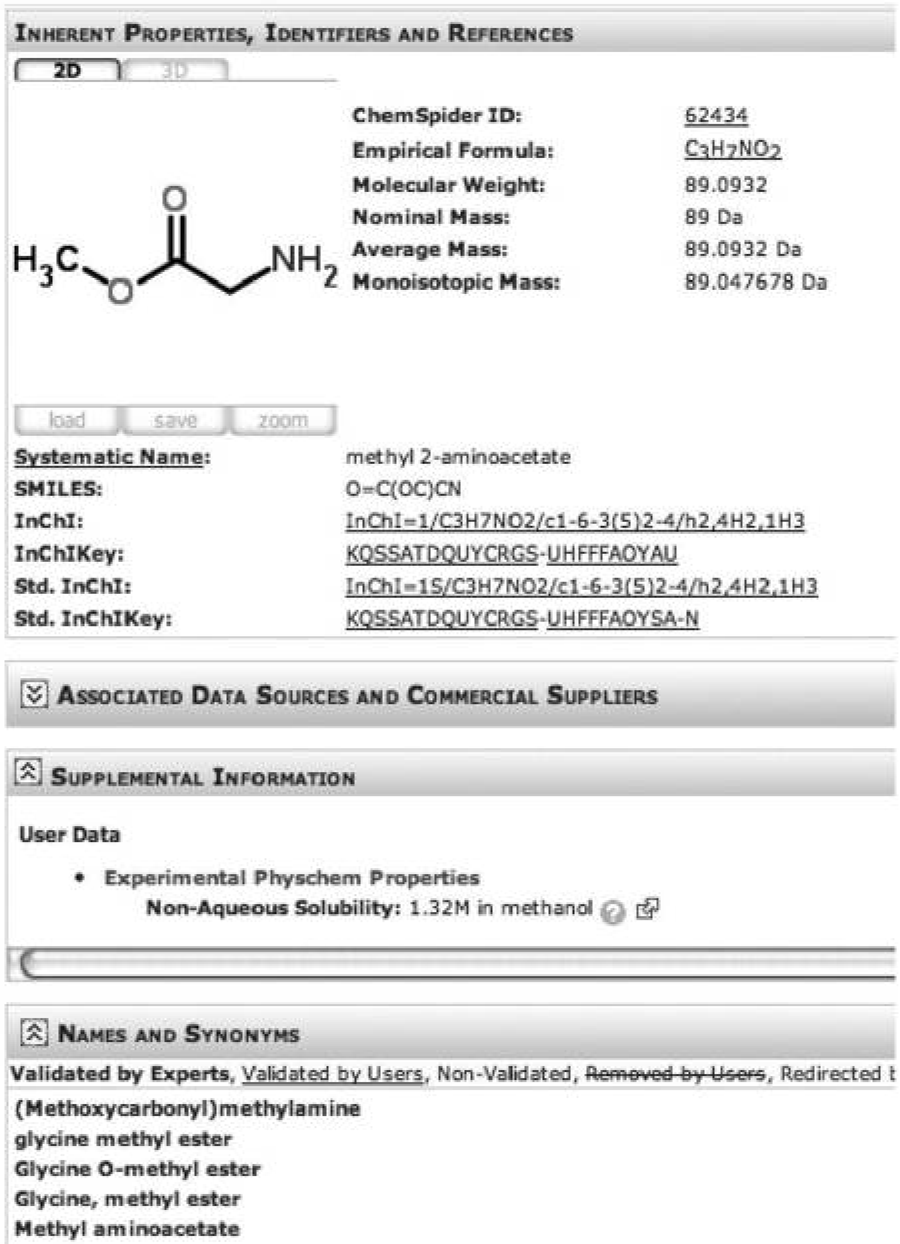
(htp[illegible]//www.chemspider.com/DataSources.aspx)[illegible]

ChemSpider[illegible][4][illegible]ChemSpider[illegible]

[illegible]ChemSpider[illegible][5][illegible]

ChemSpider[illegible]ChemSpider[illegible]ChemSpider[illegible](htp[illegible]//www.chemspider.com/blog/care-in-nomenclature-handling-andwhy-visual-inspection-will-remain.html)[illegible]

[illegible]ONS-Solubility[6][illegible]ChemSpider[illegible]ChemSpider[illegible]16-3[illegible]ChemSpider[illegible]



□ 16-3□□□ChemSpider□□□□□□□□□□□□□□□□□□□□55□

□□□GDoc□□

□□
□□
□□□□□□□□□□□□□□□□□□

□□□□□□□□□□□□□ChemSpider□□ChemSpider□□□□□□□□□□□□□□ChemSpider□□□□SQL Server□□□□□□□ ChemSpidr□□Web□□

□□□□□□□□□□□□□□□□□

□□G□□Doc API□□□GDoc□□SMILES□□□ChemSpider□□□

□□□□□(Te Resource Description Framework,RDF)□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□“□□”□□“□□”□□“□”□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

solute#59[illegible]URL[illegible]RDF[illegible]“[illegible]”[illegible]
[illegible]“[illegible]”[illegible]RDF
[illegible]RDF[illegible]“[illegible]……[illegible]”[illegible]“[illegible]”[illegible]
[illegible]ons[illegible]
[illegible]Dublin Core(dc)[illegible]
[illegible]chem[illegible]
http[illegible]//rdf.openmolecules.net(RON)[illegible]
[illegible]

[illegible]G[illegible]Doc API[illegible]URL[illegible]RDF[illegible]“2-[illegible]”[illegible]“2-[illegible]”[illegible]RON[illegible]RON[illegible]HTTP URL[illegible]URL[illegible]RON[illegible]RDF[illegible]SMILES[illegible]InChI[illegible]SMILES[illegible]

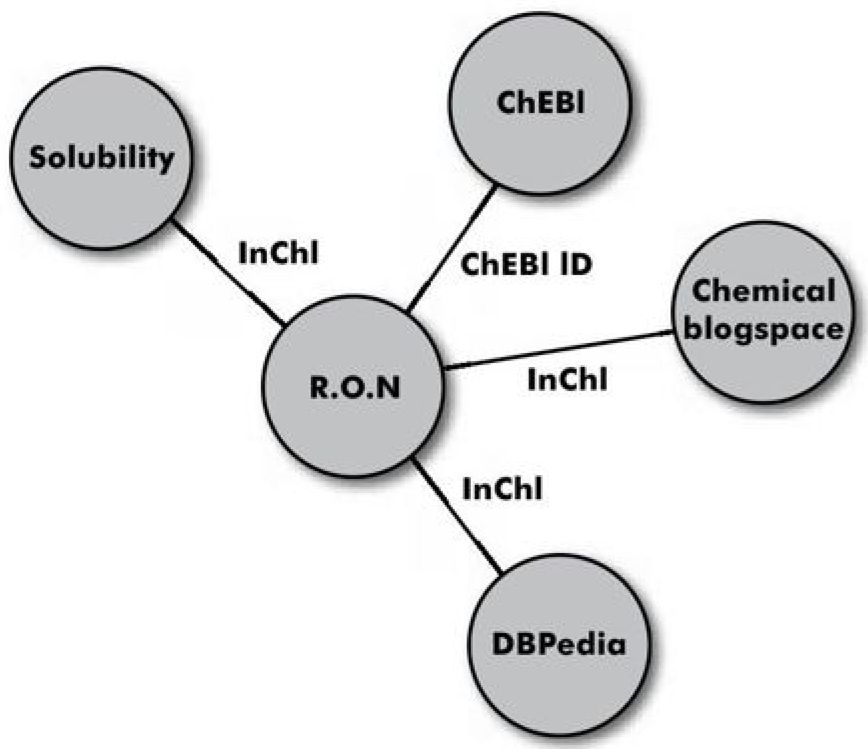
```
□ons:Solute rdf:about="http□//spreadsheet.google.com/....../onto#solute59"□
□rdfs:isDefinedBy rdf:resource="http□//rdf.openmolecules.net/?InChI=
1/C8H14O2/c1-2-3-4-5-6-7-8□9□10/h6-7H□2-5H2□1H3□□H□9□10□"/□
□chem:inchi□InChI=1/C8H14O2/c1-2-3-4-5-6-7-8□9□10/h6-7H□2-5H2□1H3□
□H□9□10□□/chem:inchi□
□chem:smiles□CCCCCC=CC□=O□O□/chem:smiles□
□dc:title□2-octenoic acid□/dc:title□
□/ons:Solute□
```

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□RDF□□□□□RDF
□□□
□□□ID□□□□□□□□□□□□□□□□□□□□□RDF□□□□□□□□□□□RDF□□□□
XML□□ons□□
http□//spreadsheet.google.com/plwwufp30hfq0udnEmRD1a
Q/onto#□□□□□XML□□□□□□(masurement179□□□□□□□
□“measurement179”□□□URL□□

```
□ons:Measurement RDF:about="measurement179"□
□ons:solubility□0.44244235315106□/ons:solubility□
□ons:solvent RDF:resource="solvent8"/□
□ons:solute RDF:resource="solute26"/□
□ons:experiment RDF:resource="experiment2"/□
□/ons:Measurement□
```

□□□□□□□□□□□[7]□□□□□RDF□□□□□SPARQL□□□□□□□□□□□□□□□
□□□
□□(htp□//github.com/egonw/onssolubility/tree/□□□□□□□RDF□
□□□□□
http□//github.com/egonw/onssolubility/tree/master/ons.solu
bility.RDF/ons.RDF□□□□□□□□□□(Te Chemistry Development
Kit,CDK□□□http□//cdk.sourceforge.net/□□□□SMILES□□□□□
InChI□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□RDF□□□□□□□□□□□
□□□
□□□□□□□□□□□□□□□□□□□□□

RDF[illegible]16-4[illegible]DBPedia[illegible]DBPedia(http[illegible]//dbpedia.org)[illegible]RDF[illegible]DBPedia[illegible](Smple Knowledge Organization System,SKOS[illegible]http[illegible]//www.w3.org/TR/skos-primer/[illegible]“[illegible]”[illegible]DBPedia[illegible]DBPedia[illegible]RDF[illegible]RON[illegible]ID)[illegible]DBPedia[illegible]DBPedia[illegible]InChI[illegible]DBPedia[illegible]ONS[illegible]RDF[illegible]



□ 16-4 □□□R D F□□□□□□□□□□□□□□□□□□□□□RO N(http □ //rdf.openmolecules.net) □ □ DBPedia □ Chemical Blogspace□ChEBI□□□□□□□□□□□□□□□□□□□

□□□□□□□□□□□□□□Web□□□□□□□□□□□□□□RON□□RDF□□□□□□□□□□□□RDF□□□□Chemical blogspace(http□//cb.openmolecules.net/□□□□□□□□□□□□□□InChI□URI□□□□□□□rdf.openmolecules.net□□□□□□□□□□□□□□□□□□□□□□□□□□□RDF□□

[1]CAS□□□□□□□□□□□□□□□□CAS□□□108-88-3□

[2]□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

[3]IUPAC□□□□□□□□□□

[4]□link farm□□□□□□□□□□□□□□

[5]□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

[6]□Open Notebook Science-Solubility□□□□□□□□□□

[7]□□□“□□”□“□□”□“□”□RDF□□□

[illegible]

[illegible]

[illegible][1][illegible]ID[illegible](mshup)[illegible]CDK(Steinbeck 2006[illegible]REST[illegible]http[illegible]//www.chembiogrid.org/cheminfo/rest/desc/descriptors/c1ccccc1COCC[illegible]URL[illegible]URL[illegible]XML[illegible]URL[illegible]XML[illegible](M)[illegible](AogP[illegible](TSA)[illegible]Web[illegible]

[illegible]Web[illegible]GDoc[illegible]Indiana University[illegible]Web[illegible]

[illegible]

(htp[illegible]//oru.edu/cccda/sl/descriptorspace/ds.php)[illegible]16-5[illegible]X

[illegible]Y[illegible]

[illegible]

[illegible]A[illegible]ChemSpider[illegible]

[illegible]

[illegible]Second Life[2]

(htp[illegible]//www.secondlife.com[illegible]16-6[illegible]3D[illegible]GDoc[illegible]

[illegible]Second Life[illegible]

[illegible]

[illegible]Second Life[illegible]

Second Life[illegible]

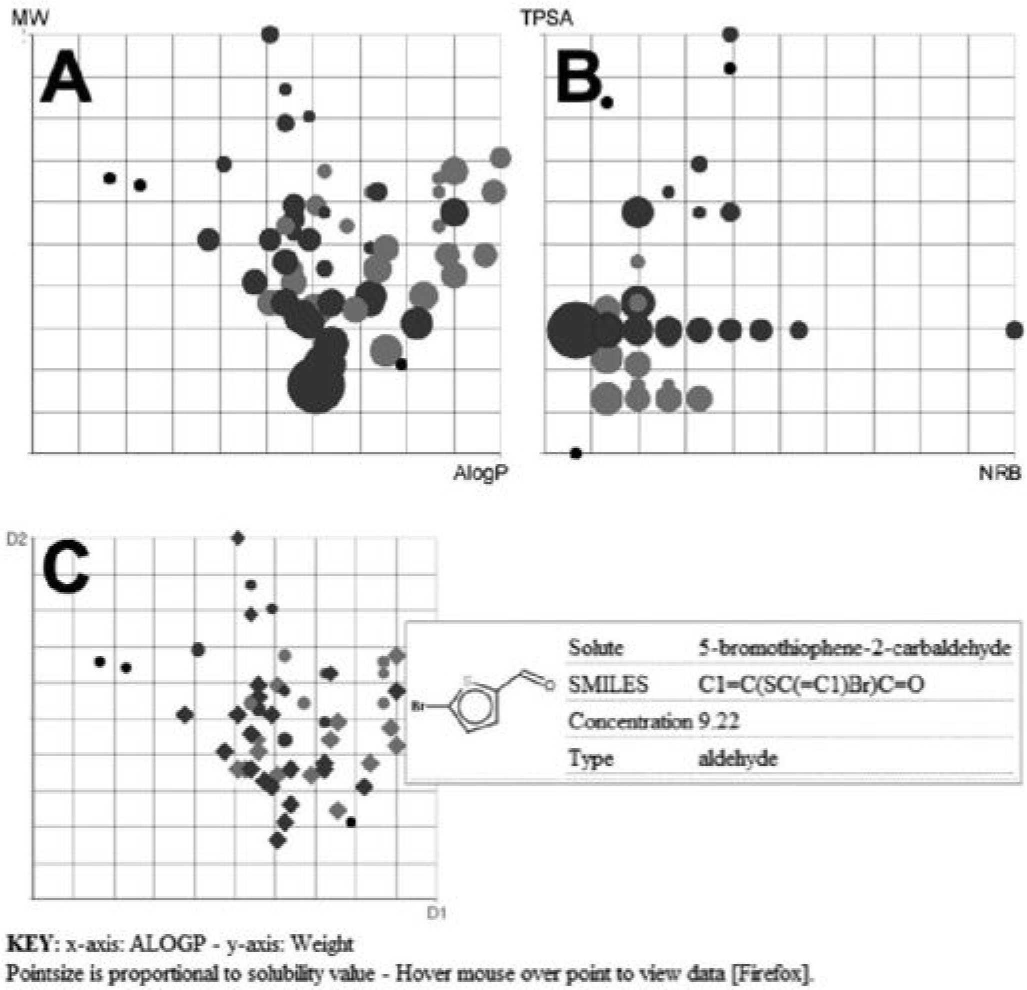
[illegible]Second Life[illegible]

Web[illegible]Web[illegible]

[illegible]

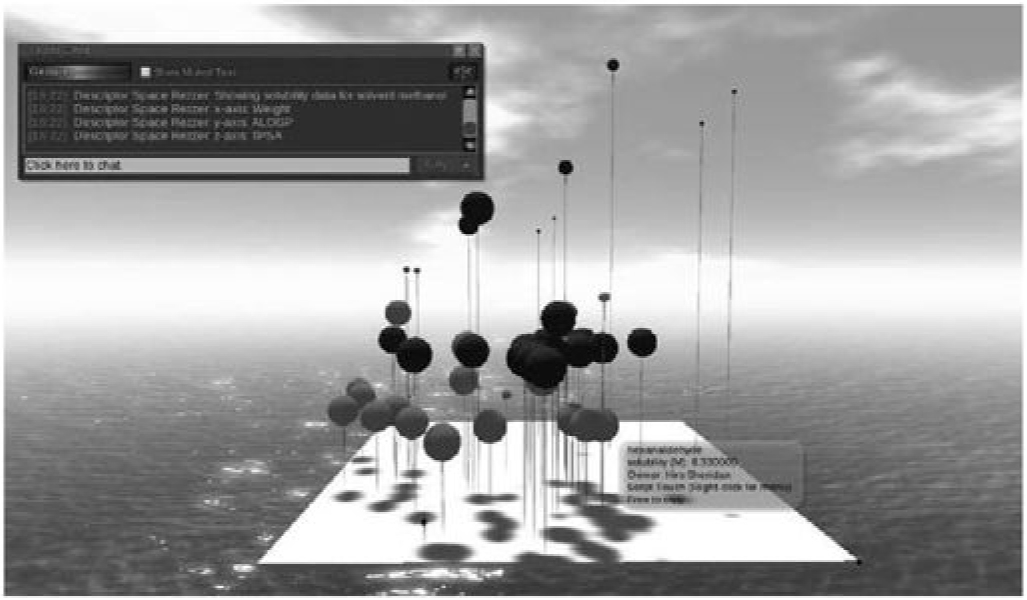
[illegible]Second Life[illegible]

[illegible]



[illegible] 16-5[illegible]A[illegible]B[illegible]C[illegible]56[illegible]

[illegible]SMILES[illegible]Web[illegible]RDF[illegible]



□ 16-6□□□Second Life□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□16-5□□□□□□□□□□□□□□□□□□□□□□□□□□□□□http□//slurl.com/secondlife/Drexel/165/178/24□□□□Second Life□□Drexel□□□□□57□

□□RDF□□□Web□□

[1]□Chemical Space□□□□□□□□□□□□□□

[2]□□□□□□Linden□□□□□□□□□□□□□□□□□

[illegible]

[illegible]

[illegible]“[illegible]”[illegible]

[illegible]GDoc[illegible]

[illegible]feed[illegible]

[illegible]G[illegible]Doc API[illegible]

[illegible]RDF[illegible]

[illegible]ChemSpider[illegible]wiki[illegible]JavaScript API[illegible]Second Life[illegible]

[illegible](RF)[illegible]

[illegible]XML[illegible]
RDF[illegible]

[illegible]

[illegible]

[illegible]

□□

□□□□□□□□□□□□□□□□Khalid Mirza□Jennifer Hale□Tim Bohinski□□□□□□□□□□□□□□Bill Hooker□□□□□□□□□□□□□□Submeta□Nature□□□□□□□□□□□□□□□□□

□□□□

[1]Bradley,Jean-Claude. 2007."Open Notebook Science Using Blogs and Wikis."Available from Nature Precedings,http□//dx.doi.org/10.1038/npre.2007.39.1.

[2]Bradley,Jean-Claude et al. 2008."Optimization of the Ugi reaction using parallel synthesis and automated liquid handling."Journal of Visualized Experiments,http□//dx.doi.org/10.3791/942.

[3]Steinbeck,C. et al.2006.Current Pharmaceutical Design□12□2110-2120.

□17□　□□□□□□□□□□□□□□□□

Brendan O'Connor□Lukas Biewald

[illegible]

[illegible]Chris Van Pelt[illegible]FaceStat.com[illegible]17-1[illegible]10[illegible]1000[illegible]

·[illegible]

·[illegible]

·[illegible]

·[illegible]

[illegible]“[illegible]Hot-or-Not”[illegible]

[illegible](sereotype)[illegible]



[illegible] 17-1[illegible]FaceStat[illegible]58[illegible]

[illegible]——[illegible]

□□□□□

□□□□□□□□□□□□□□□FaceStat□□□□□SQL□□□□□□□□□□□□□□□□□□□□□□□□□□□(fce ID,attribute,judgment)□□□□□□□□□□□□□□□□□□□□□1000□□□□□□□□□□□□□□□□

```
face_id key value
149777 describe serious
18717 trustworthy 3
140467 attractive 2
149777 describe five-head
……
```

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□“□□□□□□□□”□□□□□□□□□□□□□□□□□Unix□□

```
Look at$cat data.tsv|
age judgments'grep"age"|
values cut-f3|
and count how many times sort|
each value occurs,uniq-c|
and order by this count.sort-nr
```

□□□□□shell□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□——□□□□□□□“□□□□□□□□”□Web□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

```
7047219
7002122
6938718
```

```
6842317
……
2724\r\n
2717\r\n
2301
2116\r\n
……
1 old enough to know better
1 hopefully over 21
1 e
1??
……
```

FaceStat[illegible]8[illegible]\r\n[illegible]bug[illegible]sort|uniq-c|sort-nr[illegible]bug[illegible]

[illegible]100[illegible]“mr.cool”[illegible]

[illegible]1[illegible]“[illegible]”[illegible]“[illegible]”[illegible]2[illegible]17-1[illegible]

表17-1：每张脸的数据

男性	年龄	智商	魅力	警察关系
TRUE	24.26667	NA	2.800000	NA
TRUE	47.00000	3.400000	2.120000	3.2
TRUE	29.27273	2.700000	2.083333	1.8
FALSE	17.63636	3.111111	2.428571	NA
FALSE	19.58333	NA	2.750000	NA
TRUE	22.80953	NA	2.250000	NA
TRUE	29.77778	1.833333	1.900000	NA
FALSE	18.16667	NA	2.571429	NA
TRUE	46.60000	3.200000	2.120000	3.4
TRUE	52.06667	3.000000	2.080000	NA

[illegible]20[illegible]

http[illegible]//data.doloreslabs.com[illegible]

[illegible]

[illegible]17-2[illegible]

表17-2：数据分析包的比较

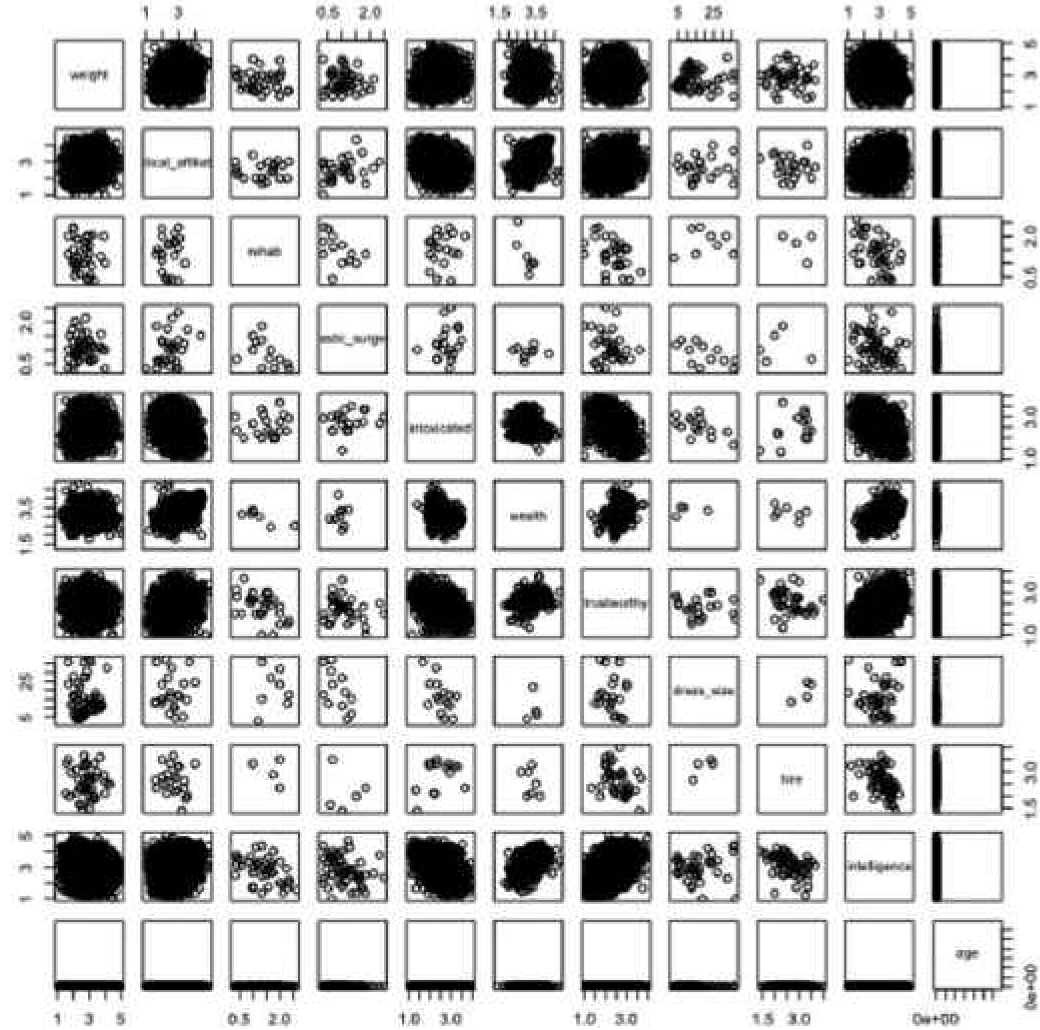
名字	优点	缺点	是否开源	典型用户
R	库支持，可视化	很陡的学习曲线	是	统计学
Matlab	优雅的矩阵支持，可视化	代价高，统计支持不全	否	工程
SciPy/NumPy/Matplotlib	Python：灵活通用的编程语言	各个组件集成很差	是	工程
Excel	简单、可视、灵活	大数据集、弱数值和编程支持	否	商业
SAS	很大的数据集	非常复杂，学习代价最高	否	商业
SPSS，Stata	简单的统计分析	不灵活	否	科学（生物学和社会学）

[illegible]R[illegible]R[illegible]http[illegible]//www.r-project.org[illegible]

R[illegible]

```
    Load the data[illegible]data=read.delim
[illegible]"http[illegible]//data.doloreslabs.com/face_scores.tsv"[illegible]sep="\t"[illegible]
    and plot.[illegible]plot(data)
```

[illegible]R[illegible]17-2[illegible]——[illegible]



□ 17-2□□□□□□□□□□□□

□□□□□□□□□□□□□□□□□□□□□□□□17-3□□

```
□hist(data$age)
```

Select records with age greater than 100. `> data[which(data$age > 100),]`

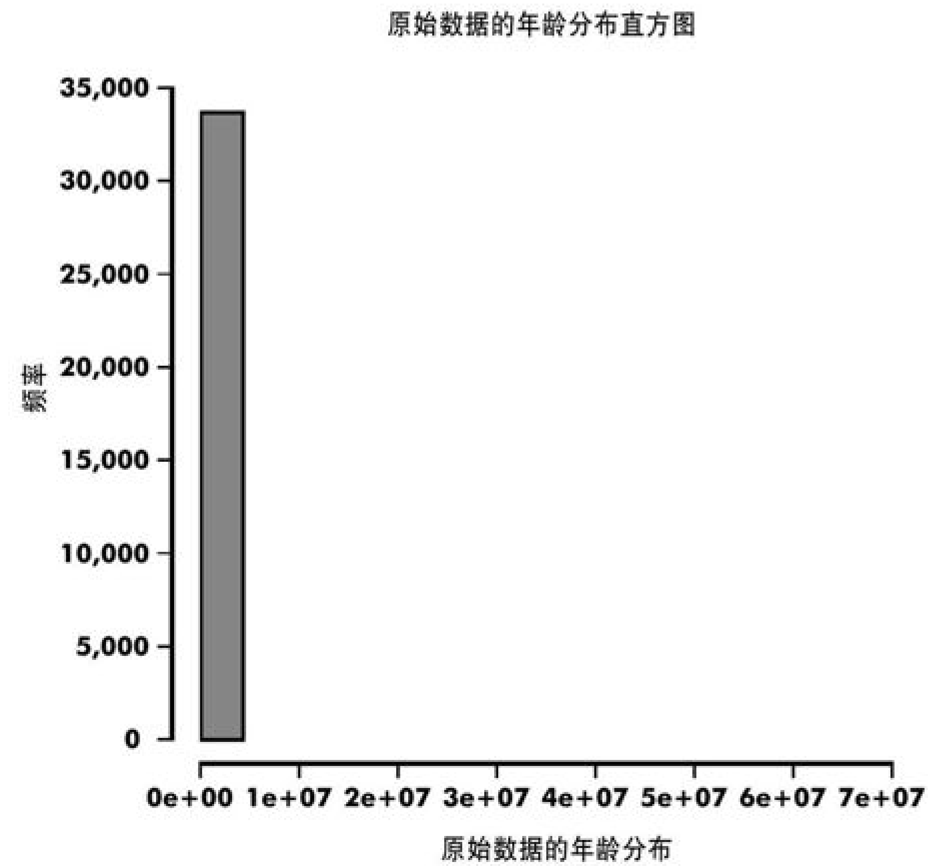
id	评价数量	年龄	性别	魅力	智商
40623	150	402.3333	TRUE	2.416667	NA
57021	133	47882.3010	TRUE	NA	NA
66441	197	66666692.0000	TRUE	NA	NA

□ 17-3□□□□□□□□□□□□□

□□□□□□□□□□□□□□□□x□□□□□□□□7000□□□□□□□□□□□□□□□□□□□

□□

Select records with age greater than 100. □data[which(data$age□100□□]



□□□R□□□□□□□□□□□□□□□□□□□□□□□□□□

□□

Subselect rows with age less than 100. □clean_data=data[which(data$age□100□□]

清除无效的数据后，年龄分布如图17-4所示，这个分布就合理多了，大多数用户的年龄在18～30岁，这也合乎我们的预期。

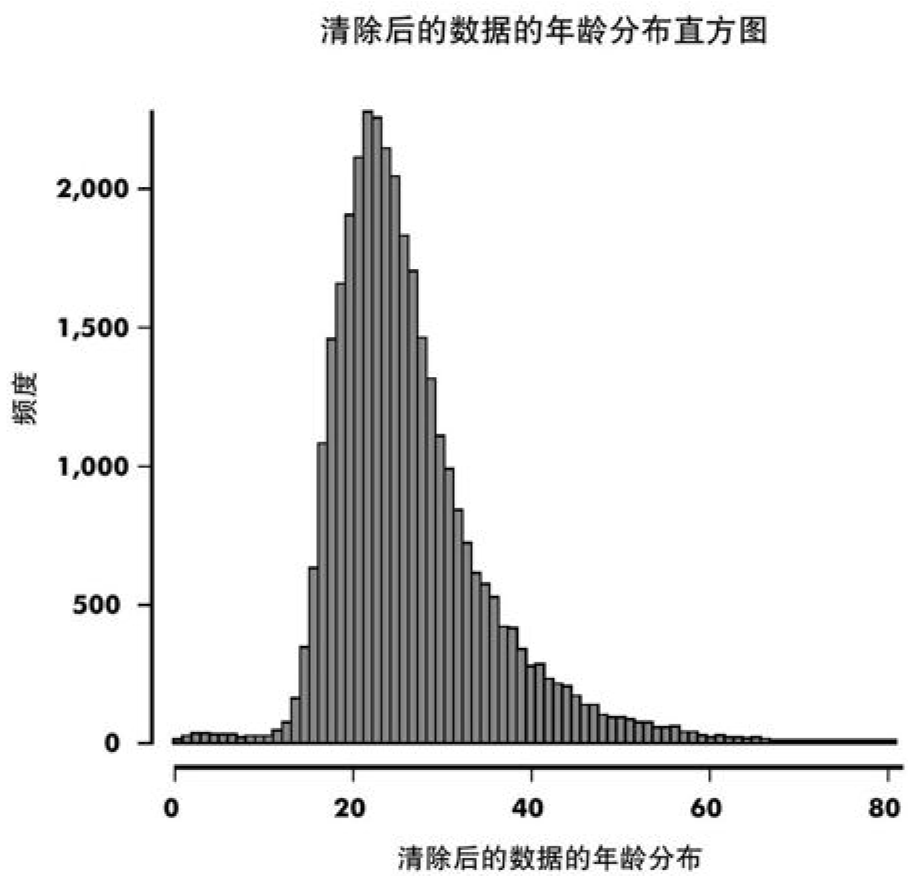


图　17-4　清除无效数据后用户的年龄分布直方图

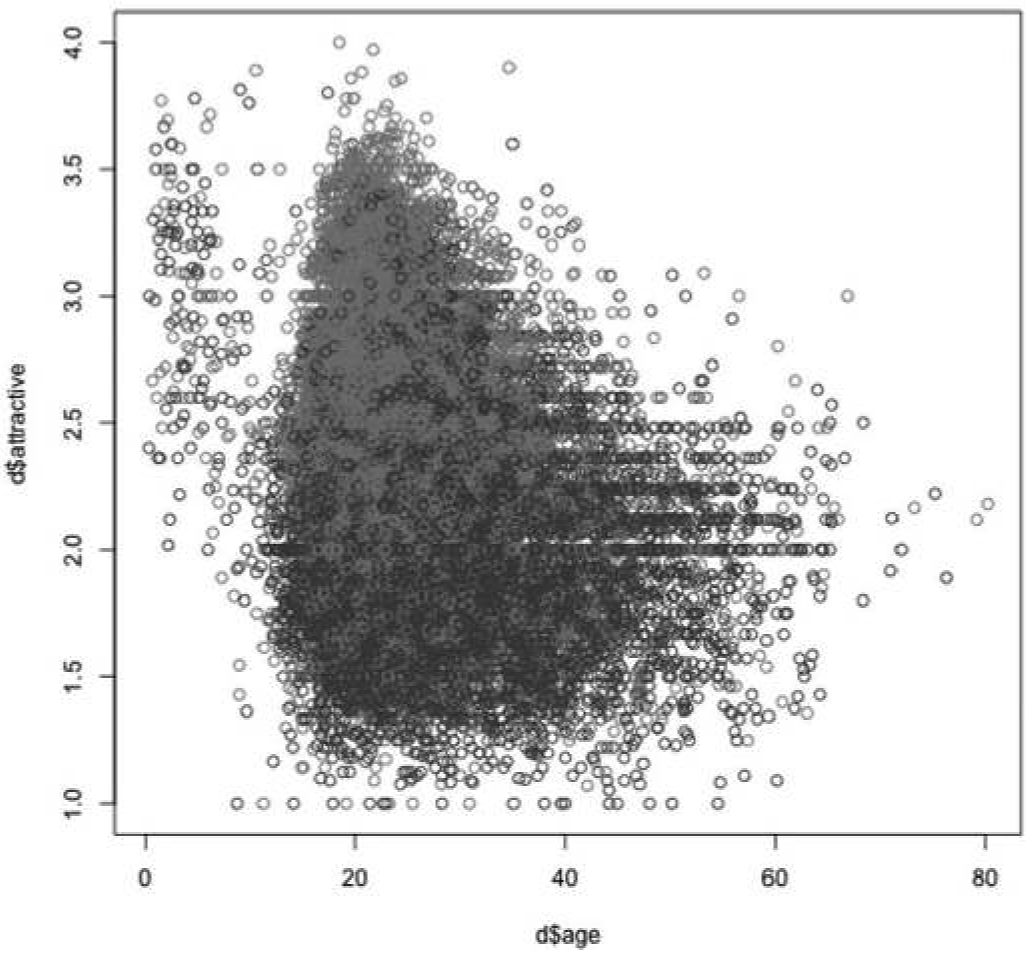
[illegible]

[illegible]17-5[illegible]

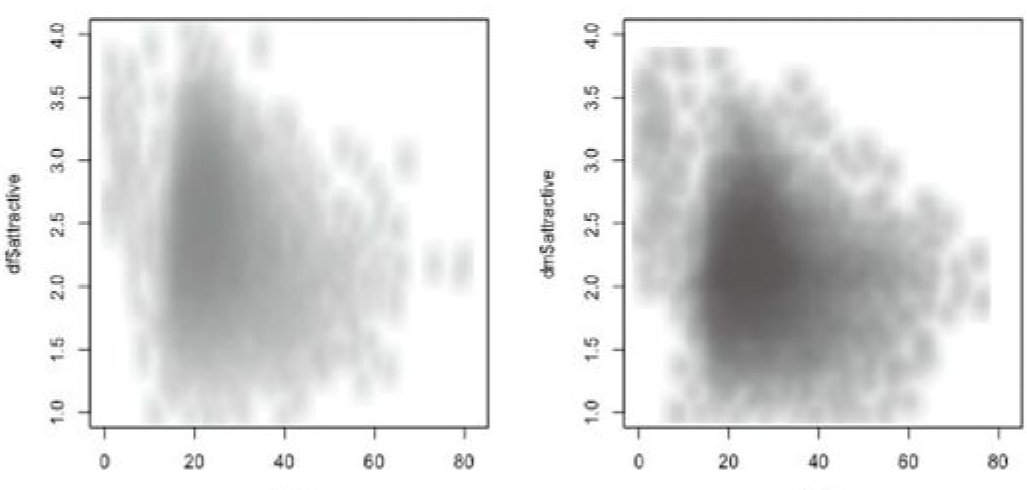
```
Draw a scatterplot of age vs.attractiveness□□plot(d$age,d$attractive□
using gender to define the
points'colors.col=ifelse(d$male□'blue'□'deeppink'□□
```

[illegible]17-6[illegible](krnel density estimation)[illegible]

```
Lay out side-by-side plots.□par(mfrow=c□1□2□□
For males and females□□dm=d[d$male□]□df=d[d$female□]
draw smoothed plots□□smoothScatter(df$age,df$attractive□
with a color gradient,colramp=colorRampPalette(c□"white"□"deeppink"□□□
and aligned axes.ylim=c□0□4□□
□smoothScatter(dm$age,dm$attractive□
colramp=colorRampPalette(c□"white"□"blue"□□□ylim=c□0□4□□
```



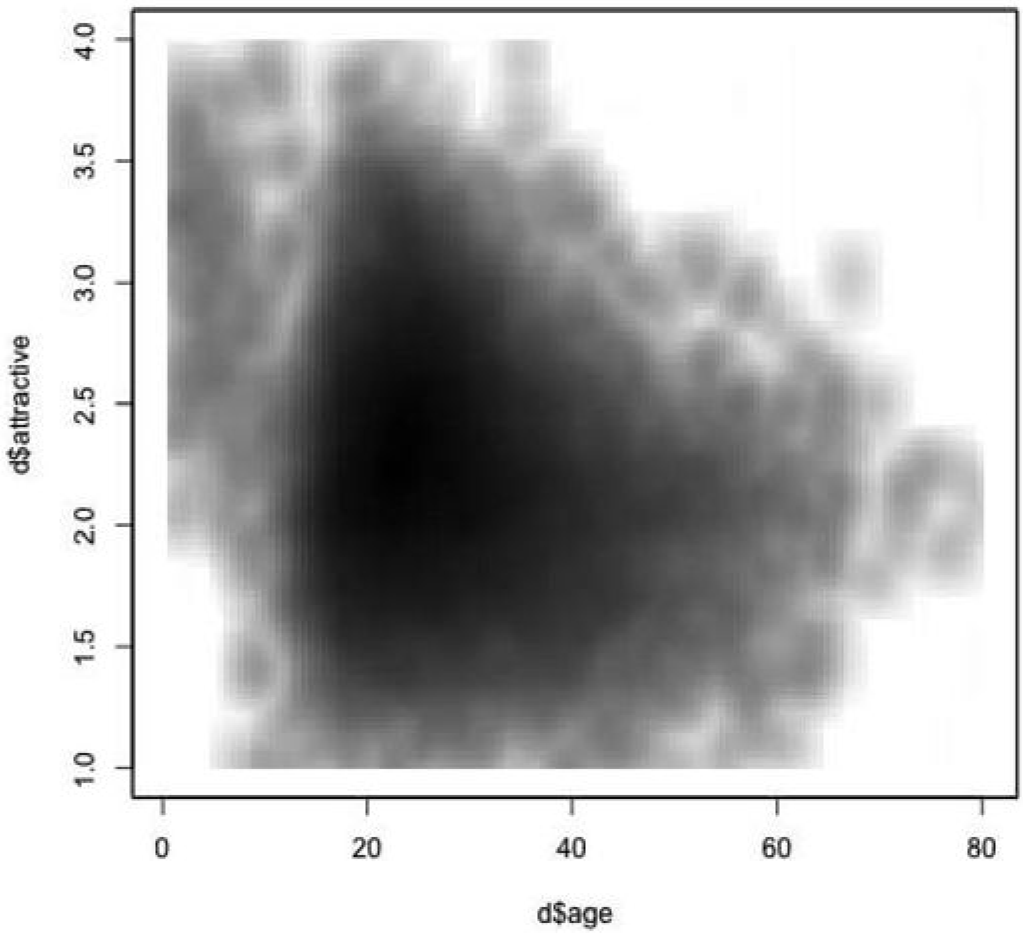
[illegible] 17-5[illegible]59[illegible]



[illegible] 17-6[illegible]60[illegible]

□□□□□□□□□□□□□□□□□□□□□17-7□□

```
□
smoothScatterMult(d$age,d$attractive,d$male,blendFun=bl_burn,colramps=
    c(colorRampPalette(c□"white"□"red"□□□colorRampPalette(c
□"white"□"blue"□□□
    colorRampPalette(c□"white"□"green"□□□□pch=""□nrpoints=10000□
```



□ 17-7□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□61□

□□□

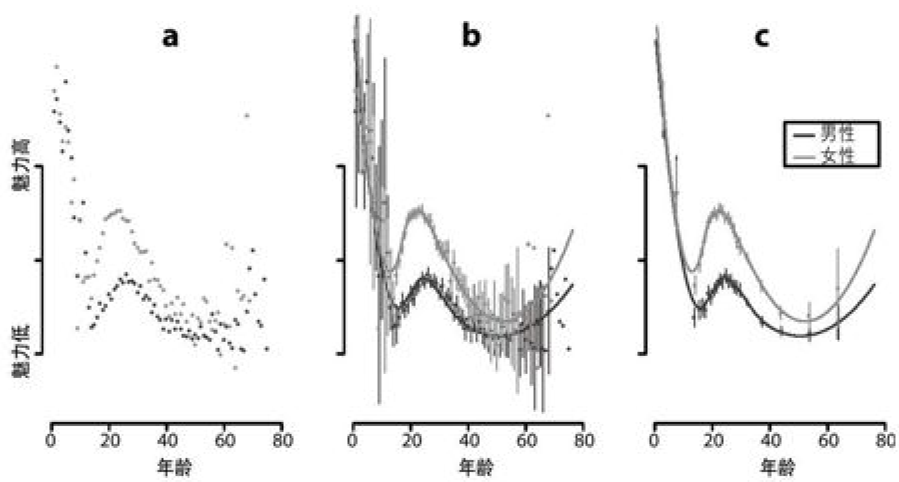
□□□□□□□□□□□□□□□□□□17-8a□□

```
For males  □dm=d[which(d$male)□]
and females□  □df=d[which(d$female)□]
average across faces  □male_avg_by_year=by(dm$attractive□
within bins  cut(dm$age,breaks=0□80□□mean)
```

```
(oe per year)   □female_avg_by_year=by(df$attractive□
then   cut(df$age,breaks=0□80□□mean)
plot them   □plot(male_avg_by_year,col='blue'□
all together.   □points(female_avg_by_year,col='deeppink'□
```

□□□

□□□□□□□□□□□□17-8b□□□□□□□□95%□□(lcally weighted regression,LOESS)□□□17-4□□□□□□□□□□——□□□□50□□□□□□□□1.7%□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□——□□□□□□□——□□□□□5□□10□□□□□□□□□□□□□□□□□□□□17-8c□□

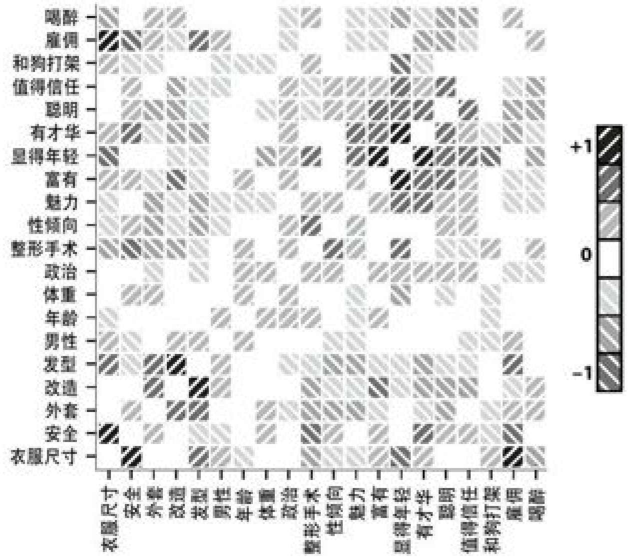


□ 17-8□□□□□□□□□□□□□□□□□a□□□□□□□□□□□□b□□□
□95%□□□□□□□□□□□□□c□□□□□□□□□□□□□62□

□□□18□□□□□□□□□□□□“□□□”□□□□□□□□□□□□□□□□27□□□□□□□□□□□□□□□□□□□50□□50□□□□□□□□□□□□□□□□□□□□□□□□□50□□□□□□□

□□□□20□□□□□□□□□□□□□□□□□□□□□□□□□□17-8□□□17-9□□

```
    First compute pairwise correlations□   □cors=cor(d,use='pair'□
    and order the attributes to try to   □ord=order.hclust(cors)
    put similar attributes next to each other.   □cors=cors[ord,ord]
    Plot the correlation matrix□   □image(cors,col=col.corrgram□7□□
    with axis labels.   □axis□1□at=seq□0□1□length=nrow(cors)□□
labels=row.names(cors)□
```



问题文本

- 衣服尺寸：我的衣服尺寸是多少？
- 安全：如果你是一名机场安检巡警，你会检查我吗？
- 外套：你喜欢我的外套吗？
- 改造：我是否需要改造？
- 发型：你喜欢我的发型吗？
- 年龄：我多大了？
- 体重：我有多重？
- 政治面貌：我的政治面貌是什么？（越高越保守）
- 整形手术：我是否做过整形手术？
- 性倾向：我的性倾向是什么？（值越高越有可能是同性恋）
- 魅力：我有多大魅力？
- 富有：我有多富有？
- 显得年轻：我是否（或者将来是否会）显得年轻？
- 有才华：我是否有才华？
- 聪明：我有多聪明？
- 值得信任：我值得信任的程度有多高？
- 和狗打架：你觉得我是否能打过一只中等大小的狗？
- 雇佣：你会雇佣我吗？
- 喝醉：我有多醉？

图　17-9□Pearson□□□□□□□□□□□□□□□□□□□□□□□□

□□□□□□□□□□63□

□□□□□□□□□□□□□□□□□□□□

·□□□□□□□□□

·□□□□□□□□□□□□□

· [illegible]

· [illegible]

· [illegible]

[illegible]“[illegible]·[illegible]”[illegible]

[illegible]

[illegible]“[illegible]”[illegible]“[illegible]”[illegible]“[illegible]”[illegible]“[illegible]”[illegible]“[illegible]”[illegible]“[illegible]”[illegible]“[illegible]”[illegible]“[illegible]”[illegible]“Nancy[illegible]”[illegible]“[illegible]”[illegible]“......”[illegible]“plokmnjiuhbygvt fcrdxeszwaq”[illegible]

```
Load our tags  [illegible]face_tags=read.delim[illegible]"face_tags.tsv"[illegible]sep="\t"[illegible]as.is=T[illegible]
then count  [illegible]counts=table(face_tags$tag)
and rank them.  [illegible]sorted_counts=sort(counts,decreasing=T[illegible]
Show the most common tags.  [illegible]sorted_counts[1[illegible]20]
```

[illegible]

cute	pretty	happy	nice	fun	young
81333	40954	36263	33221	30622	27900
sweet	friendly	cool	weird	hot	gay
20362	14895	14709	12731	12662	12409
Cute	funny	scary	sexy	old	goofy
12132	11508	11445	11287	10958	10511
emo	shy				
10292	10207				

[illegible]

```
Show the least common tags.  [illegible]tail(sorted_count[illegible]20[illegible]
```

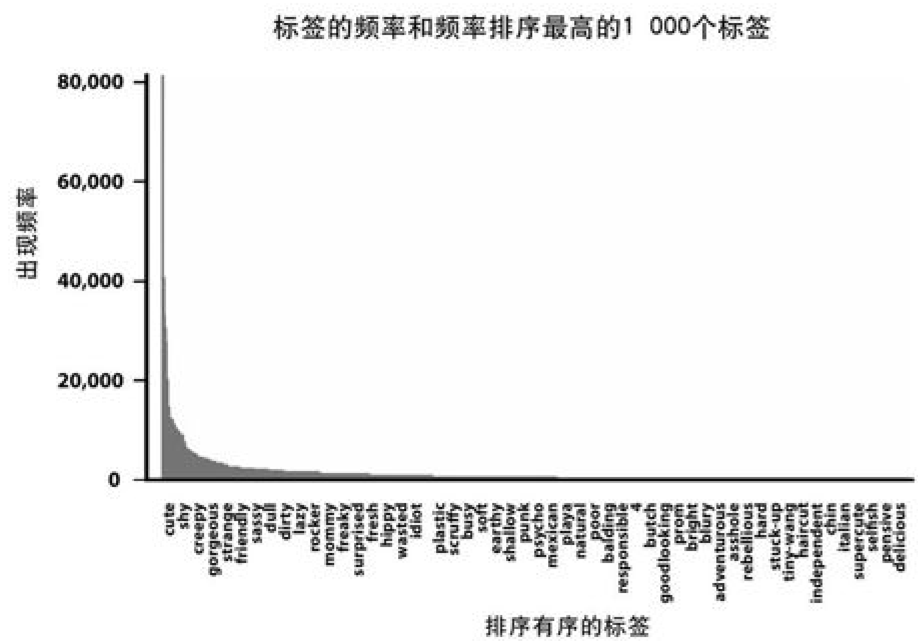
überdude	übersöt	ünsall
1	1	1
ýour.nose.is.sexymamama!!	我	浅
1	1	1
良	?	♥haiir!!
1	1	1
白人	賢母	— —;
1	1	1
шдд	オタク	ロソリー
1	1	1
ешкув	сшеет	херня
1	1	1
ダースベイダー	Красивая!	
1	1	1

□□□□□□□□□□□□□“cute”□“Cute”□□?□□□□□□□ASCII□□□□?□□“hot”□“HOT□”□□

□□240□□□□□□□29□□□□□□□□□1000□□□□□□□□□140□——□□“cute”□□□3.6□□□□□□□□□□“pretty”□□□□□□□□□□□

```
For the top 1□000 tags□   □s=sorted_counts[1□1000]
draw a plot of their counts.   □barplot(s)
```

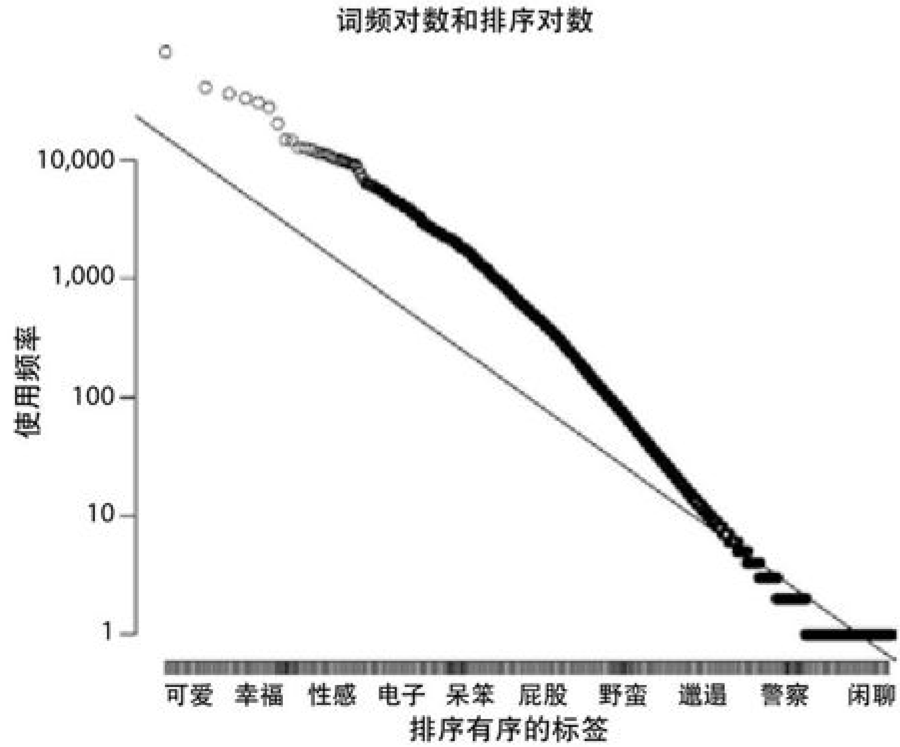
1935□□□□□□George Zipf□□□□□□□□□“□□□□□”(pwer law)□□□□n□□□□□□□□1/ns)□□□□□□s□□Nassim Nicholas Taleb□□The Black Swan□(Rndom House□□□□Chris Anderson□□The Long Tail□(Hperion□□□□□□□□□□□□“□□”□“□□”□□□□□□□□□□□□□□□□□□□□□□□□□22□□□□□□□□□□□76%□□□□□□



□ 17-10□□□□□□□□1000□□□□□□□□

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□17-11□□

```
Plot log ranks   □log_ranks=log□1□length(sorted_counts)□
against log frequency.   □plot(log_ranks,log(sorted_counts)□
```



□ 17-11□□□□□□□□□□□□□□□□□□□□□□□□

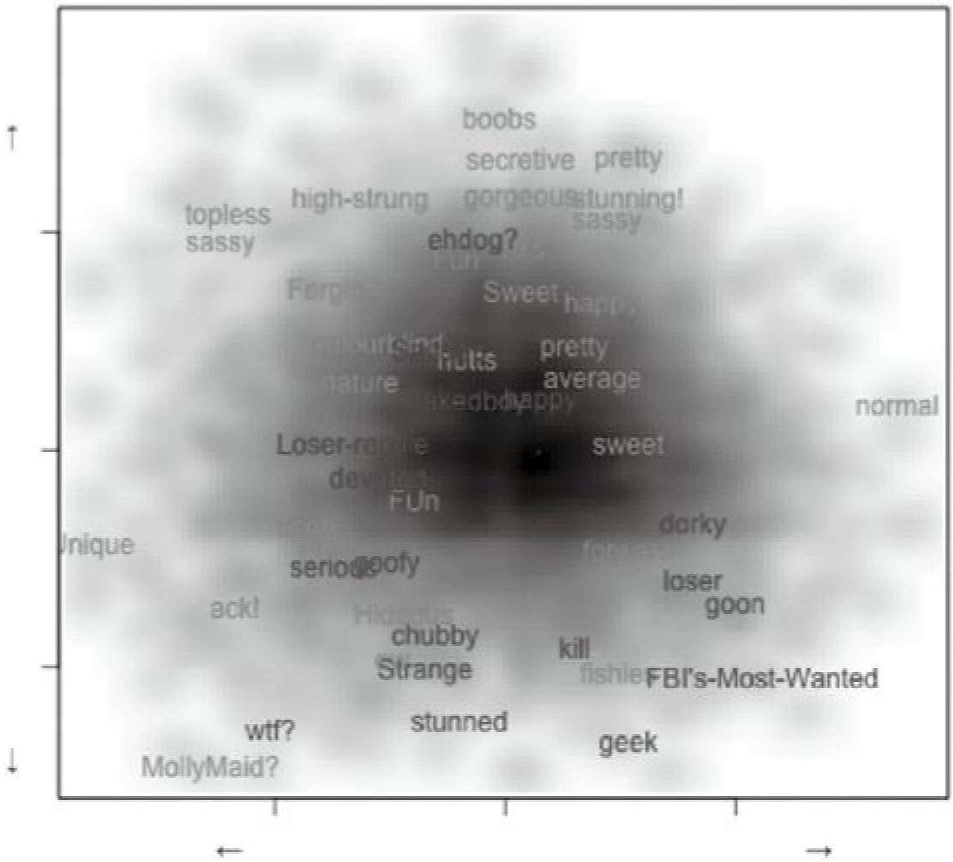
□□□□□□□□-□□□□□□□□□□□□□

Fit a model of log count against log rank □model=lm(log(sorted_counts)□log_ranks)
and draw it on Figure 17-10. □abline(model)

□□□□□□□□□□□$1/n^{80}$□□□□□□□□□□□□□□□□□□□□□□76%□□□□□□□□□□□

□□□□□□□——□□□□□Web□□□□□□□□-□□□□□□——□□□□□□□[1]□□□□FaceStat□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

□□□17-12□□



□　17-12□□□□□□□□□□□□□□□□□□□□□□□□64□

□□□“□□”□□□□□□□□□□□□□□“□□□□”□□□□□“□□/□□”□□□□□□□“□□□□□”□□□“□□/□□”□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

[1]Zipf,George.1935.□The Psychobiology of Language□(MT□□□□□http□//en.wikipedia.org/wiki/Zipf's_law□

[illegible]

[illegible]T[illegible]G[illegible]

$$\frac{\text{性别是}G\text{的脸的标签}T\text{的出现次数}}{\text{标签}T\text{出现的总次数}}$$

[illegible]1[illegible]100[illegible]

[illegible]——[illegible]Pr(G|T[illegible]——[illegible]

[illegible]

	G	T	Ratio
daddy	122	122	1.0000000
fatherly	115	115	1.0000000
fratboy	177	177	1.0000000
father	172	173	0.9942197
dad	341	343	0.9941691
douche	229	231	0.9913420
Handsome	110	111	0.9909910
scruffy	149	151	0.9867550
bald	343	350	0.9800000

（续）

	G	T	Ratio
jock	395	404	0.9777228
handsome	510	524	0.9732824
thug	141	145	0.9724138
tool	255	264	0.9659091
player	522	542	0.9630996
Gay	307	319	0.9623824
jerk	131	137	0.9562044
gamer	103	108	0.9537037
fag	148	156	0.9487179
pimp	121	128	0.9453125

[illegible]

	G	T	Ratio
Bubbly	118	118	1.0000000
Mom	161	161	1.0000000
busty	148	148	1.0000000
milf	267	267	1.0000000
mom	1,088	1,088	1.0000000
motherly	396	396	1.0000000
partygirl	221	221	1.0000000
mommy	307	308	0.9967532
mother	358	360	0.9944444
ditzy	144	145	0.9931034
fjortis	113	114	0.9912281
MILF	103	104	0.9903846
Pretty	926	935	0.9903743
cheerleader	159	161	0.9875776
boobs	153	155	0.9870968
makeup	143	145	0.9862069
bitchy	284	288	0.9861111
cougar	141	143	0.9860140
slutty	538	546	0.9853480
slut	509	517	0.9845261

[illegible]“handsome”[illegible]“gamer”[illegible]

[illegible]“Bubbly”[illegible]“slut”[illegible]

□□

□□

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□K□□□□K-means)

□□□□K□□□□□□□□□□□□□□□□□□□□□“□□”(cntroid)□□□□□□□□□□□□

□□□

□□□□□□□□□□□□□□□20□□□□□□□□□□□□□□20□□□□□□K□□□□□□□□

□□

K□□□□□□□□□□□□□□□□□□□□□□□□□——“K”□□□□□□□□□□□□□□□□

□□K□□□□□□□□

□□□□□

```
Preprocess the data□□norm_data=apply(d□2□function(x){
by changing missing values to the mean,x[is.na(x)]=mean(x,na.rm=TRUE)
and unit-normalizing values,x=□x-mean(x)□/sd(x)
which usually makes k-means work better.x}□
Then run k-means for 5 clusters□□clus=kmeans(norm_data□5□
and plot attractiveness vs.age□□plot(d$age,d$attractive□
but color by col=c□"red"□"purple"□"blue"□"orange"□
cluster assignment□"green"□"darkturquoise"□[clus$cluster]□
and have fun with unicode.pch=ifelse(d$male□'\u2642'□'\u2640'□□
```

□□

□□□

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□20□□□□□□

□□□□□□□□□□(oerlap)□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

征和结果之间的强相关性。

男性和女性在年龄、肺活量的二元特征空间中是混合的，但是这里我们采用一种监督方法——基于平滑的局部模型来预测性别。图17-13展示了一个监督分类的二元数据。测量值（肺活量）会受到男性和女性这个因素的影响，即使存在性别的混合（在其他数据集中，可能是任何类别），它们在一个小于10年的年龄窗口中男女比例的变化仍然需要考虑。

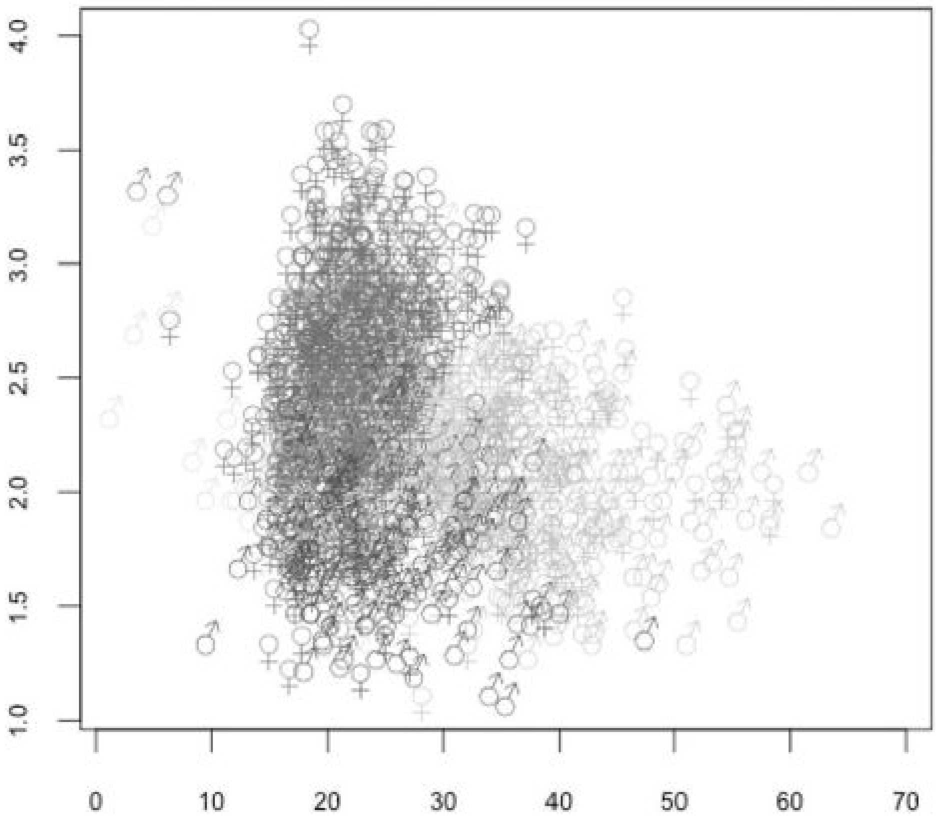
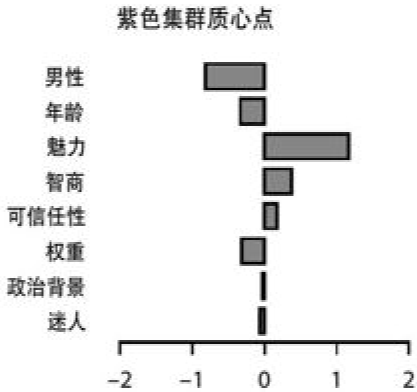


图 17-13　年龄（横轴）与肺活量（纵轴）的关系，样本数为2000，其中年龄的变化范围为65岁

正如图17-14中所显示的，这个模型将使这些变化平滑化，并且不会保证能在男性与女性的混合中进行完美的分类，但是它能很好地表示出性别对于其他变量的影响。另一种方法——最近邻法（k近邻法）就是一个很好的例子，能够说明一个小的局部模型如何能对全局模式进行很好的拟合。



紫色集群标签

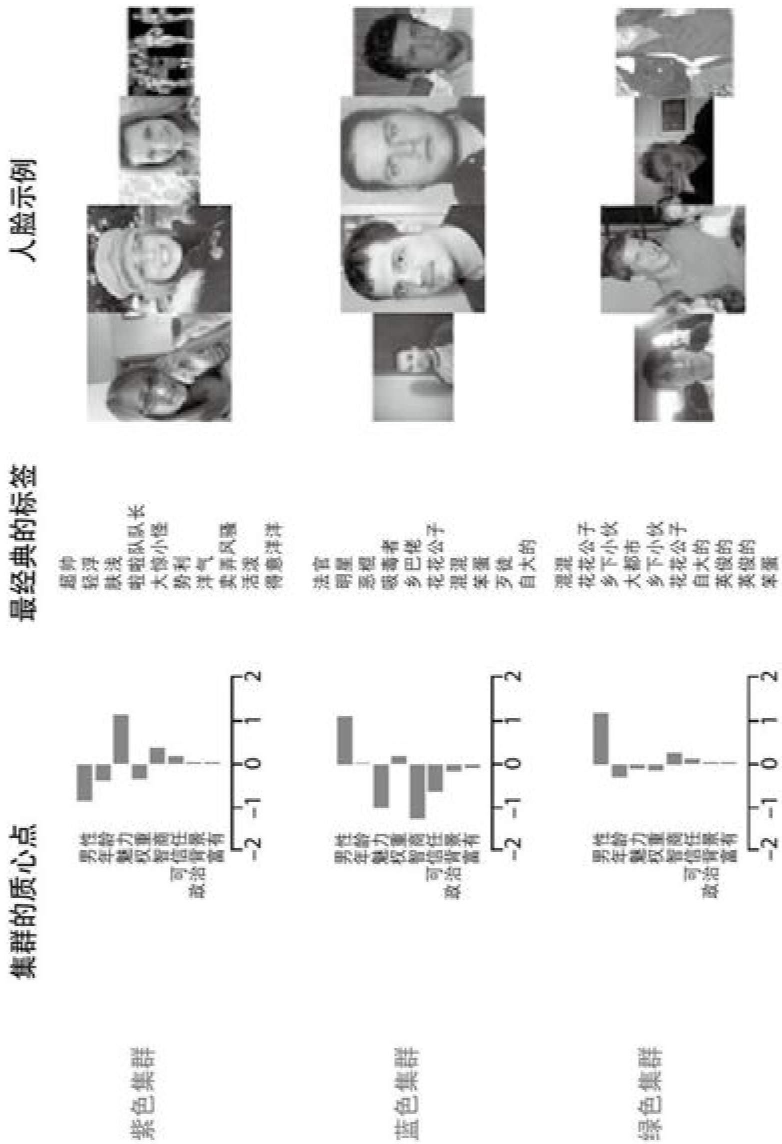
c = 紫色集群中标签的出现次数

t = 全局标签的出现次数

	c	t	ratio
超级帅	226	525	0.430
惊人	204	606	0.336
轻浮	113	439	0.257
肤浅	106	437	0.242
啦啦队队长	202	837	0.241
大惊小怪	126	561	0.224
势利	121	545	0.222
卖弄风骚	170	785	0.216
洋气	498	2327	0.214
精力充沛	123	579	0.212

[illegible] 17-14[illegible]2

[illegible]17-15[illegible]17-16[illegible]



[illegible] 17-15[illegible]66[illegible]

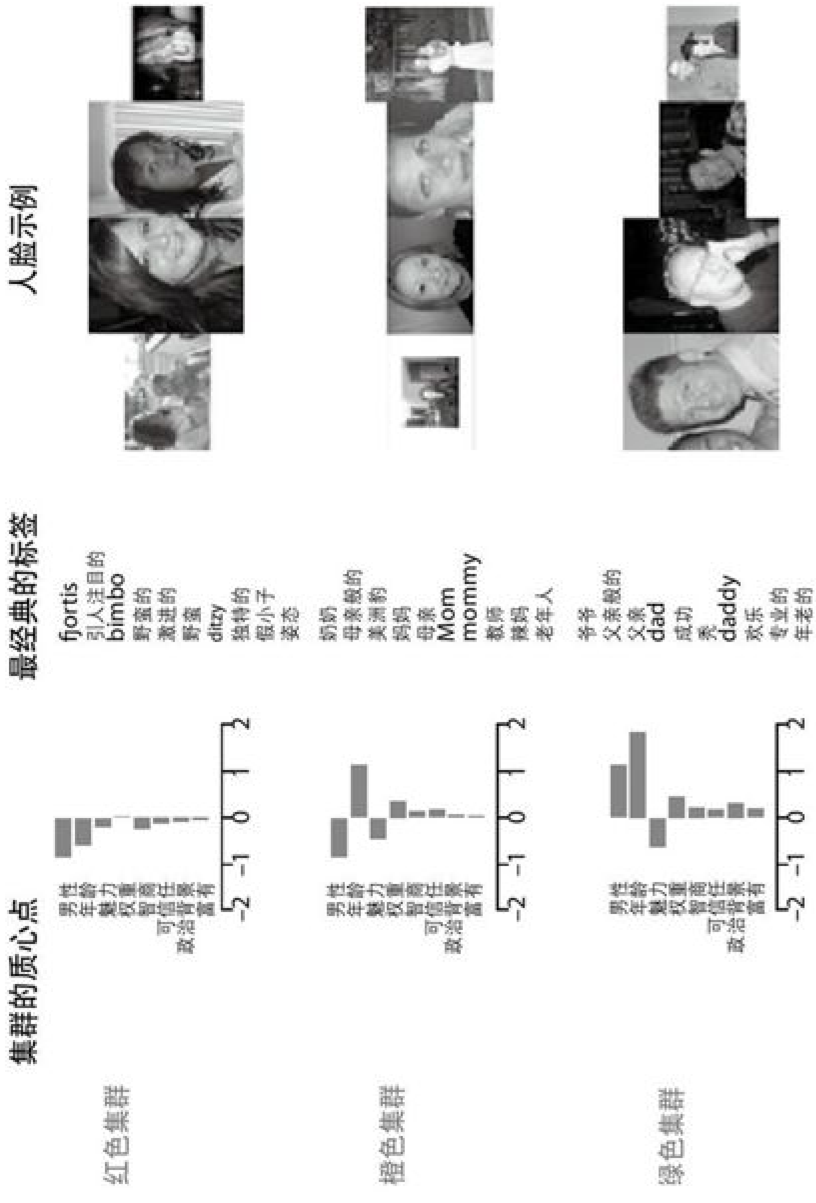
[illegible]

[illegible]

[illegible]

[illegible]

[illegible]“[illegible]”[illegible]



集群的质心点
最经典的标签
人脸示例
红色集群
橙色集群
绿色集群
男性
年龄
魅力
权重
智商
可信任
背景
富有
政治
fjortis
bimbo
ditzy
Mom
mommy
dad
daddy

图 17-16[illegible]67[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]“[illegible]”[illegible]k[illegible]

[illegible]k[illegible]

[illegible]

[illegible]“[illegible]”[illegible]

[illegible]

[illegible]Web[illegible]

□□

□□□□□□□□□□□□□□□J o a n n a G u b m a n□S a s h a G o o d m a n□J e f f Hammerbacher□Mike Love□Will Moffat□Toby Segaran□FaceStat□□□□□□□□□□□□□□Chris Van Pelt□

□□□□

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□
http□//data.doloreslabs.com□

□□□□□□□□□□R□□□□□□□□□□Web□□□

Quick-R(http□//statmethods. net)

Robert Kabacoff□□□□□□□□□□□□

RSeek(http□//rseek. org)

Sasha Goodman□□□R□□□□□□□□□□□□□□□

□□□□R□□□Web□□□□□□http□//www.r-project.org□□□□□□□□□□
□□□□□□□□□□□□□□□□□17-2□□□□□□□□□□http□//anyall.org/blog/?
p=421□□□

□□□□□□□R□□□Peter Dalgaard□□Introductory Statistics with
R□(Sringer□□□2008□□

□□□□□□R□□□□□□□□□□□□□□□□□□corrgram□flowCore□gclus□
geneplotter□plyr□pixmap□

Trevor Hastie□Robert Tibshirani□Jerome Friedman□□□□The
Elements of Statistical Learning□(Sringer□□□2008□□□□□□□□□
□□□□□□□□□□□□□□□□□□□□□□□□

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□Christopher
Manning□Hinrich Schütze□□□Foundations of Statistical

Natural Language Processing□□□□□□□(MT□□□1999□□□□

Daniel Jurafsky□James H.Martin□□□Speech and Language

Processing□(Pentice Hall□□□2008□□

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□(prtial

pooling)□□□Andrew Gelman□Jennifer Hill□□Data Analysis

Using Regression and Multilevel/Hierarchical Models□□□□□□

□□□2006□□252□258□□□□□□□□□□□“□□□□□□”(eploratory data

analysis,EDA)——□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

□□□□John Tukey□1977□□□□□□□□□□Exploratory Data

Analysis□□(Adison-Wesley□□□□□□□□□□

□□□□□□Dolores Labs□□□□□□□□□□□□□□□□□□□□□□□□□□

□□□□□□Web□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

http□//blog.doloreslabs.com□□□□□

□18□　□□□□□□□□□□□□□□□□

HadleyWickham□Deborah F. Swayne□David Poole

[illegible]

[illegible]2000[illegible]2006[illegible]

[illegible]2003[illegible]2008[illegible]50[illegible]“[illegible]”[illegible]

[illegible]R[illegible]

San Francisco Chronicle (By Area)

http //www.sfgate.com/homesales/

Web

(HML Web

rowid 1
county:Alameda County
city:Alameda
newcity 1
zip 94501
street 1220 Broadway
price $509 000
br 4
lsqft 4420
bsqft 1834
year 1910

URL http //www.sfgate.com/c/a/ year / month / day /REHS.tbl

2003/04/27 2008/11/16

URL Unix

Wget□□□□□□□□□□□□wget□□□□□□□□□□□□□□□□□□□□
□□□□□□□□□□□□□□□□□□

□□□□□□□□□□□□□□□□□□□□□□□□□□□□csv□□□□□
comma-separated values)□□csv□□□□□□□□□□□□□□
Excel□□□□□□□□□□csv□□□□□□□

```
    county,city,zip,street,price,br,lsqft,bsqft,year,date,datesold
    Alameda County,Alameda□94501□1220 Broadway□509000□4□4420□1834□
1910□2003-04-27□NA
    Alameda County,Alameda□94501□429 Fair Haven Road□504000□4□6300□
1411□1964□2003-04-
    27□NA
    Alameda County,Alameda□94501□2804 Fernside Boulevard□526000□2□
4000□1272□1941□2003-
    04-27□NA
    Alameda County,Alameda□94501□1316 Grove Street□637000□3□2700□1168□
1910□2003-04-
    27□NA
```

□□□□□□□□□□□□□□□□csv□□□□□□□□□□□□□□□□□
□c s v□□□□□□45M B□□□□□□□□□□90MB)□□□□□□□□□□
□□□□□□□□□□□□NAs□NA□□“□□□”(Nt Applicable)□□□□R□
□□□□□□□□□□□□□□□

□□□293□□□□□□□□house-sales.csv□□□□□□□□□□csv□
□□521726□□□11□□□□□□□□□□□□□□□□□□□□□□□□□□
□□□□□□□□□□□□□□□□□□□“$”□□“.”□□□□□□□□□□□□□
□□□□□□□□□□□□□□□

[illegible]

[illegible]436106[illegible]

[illegible]50[illegible]

[illegible]G[illegible]Yahoo[illegible]Web[illegible]

[illegible]1[illegible]USC WebGIS[illegible]GIS[illegible](Gldberg[illegible]Wilson 2008[illegible]8[illegible]40[illegible]5[illegible](Gldberg 2008[illegible]

[illegible]USC WebGIS[illegible]

[illegible]

[illegible]USC WebGIS[illegible]bug[illegible]http[illegible]//gpsvisualizer.com[illegible]R[illegible]

San Jose[illegible]“[illegible]”[illegible]San Jose[illegible]Berryessa[illegible]Cambrian[illegible]

[illegible]R[illegible]“[illegible]”[illegible]

[illegible]

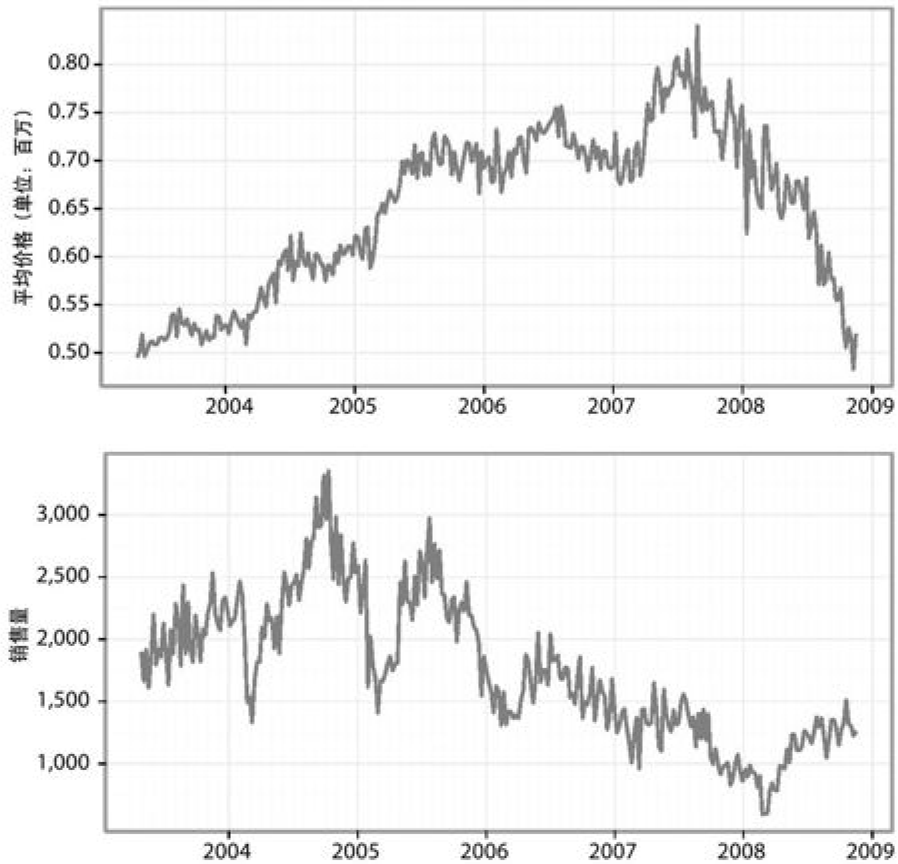
[illegible]

[illegible]18-1[illegible]293[illegible]2007[illegible]7[illegible]——[illegible]

[illegible]2004[illegible]2005[illegible]4[illegible]6[illegible]2006[illegible]2008[illegible]2008[illegible]2008[illegible]

[illegible]

□□□□□□□

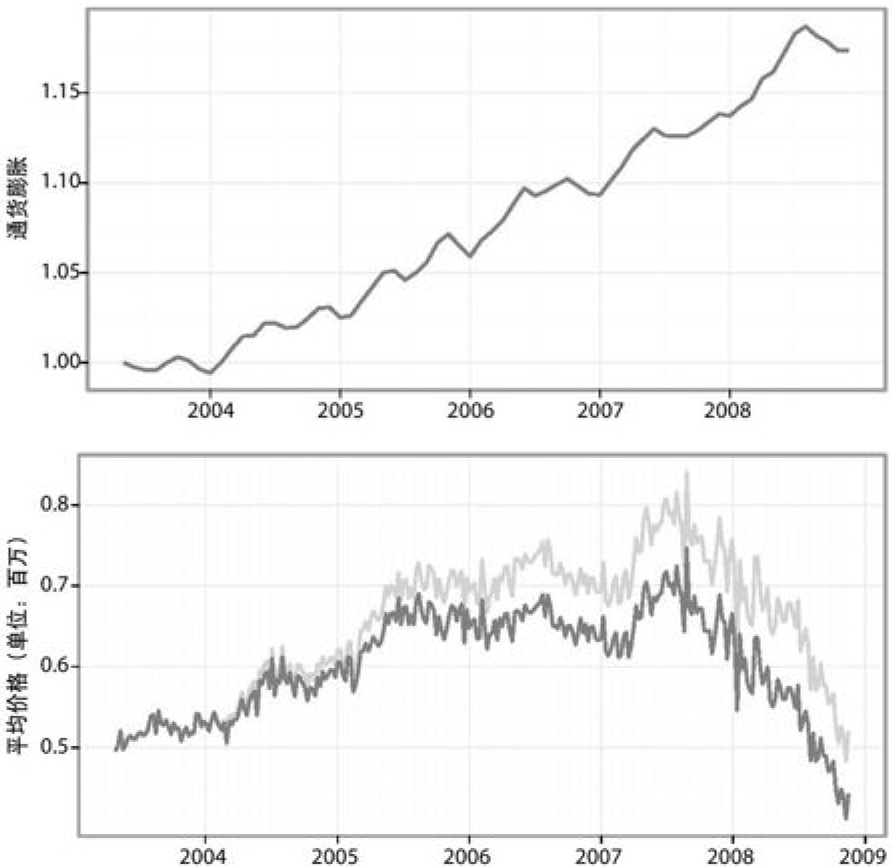


□ 18-1□□□□□□□□□□□□□□□□□□□□□□□□□□□□2008□

□□□□□

□□□□□□□□□□□□□□□□□□□□□□□□□□□□6□□□□□□ □□□□□□□□□□□□□□□□□□□□□2003□□□□□2008□□□□ □□□□□□□□□□□□□□□□□□□(Cnsumer Price Index,CPI)□□ □□□□□□□□□□□http□//www.bls.gov/CPI□□□□CPI□□□□□□□□ □□□□□□□□“□”□□□□□□□□□□□□□□□□□□(CUR0400SA0□

□□□□□□□□□□□□□□□□□□□□□□□□□2003□□□□□□□□□□□□CPI□□□□
□2003□3□□□□□□□□□□□□□□□“□□□□□□□”(Idexing)□□□□□□2003□1
□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□1.1□□□□□□□□□□□□□□
□□□□10%□□18-2□□□□CPI□□□□□□□□□□□□□□□□□□□□□□□□□□□□□
□5□□□
□□□□□□□□□□□□□□□□□□□□□CPI□□□□□□□□□□□□□□□□□□□□□□□“□
□□□□□□□”□□□□□□□□□□□□□□□CPI□□□□□□□□□□□□□□



□ 18-2□□□□□□□□□□□□□□□□□□□□□□□□1□□□□□2003□□□□□□□□□
□□

Monterey San Benito San

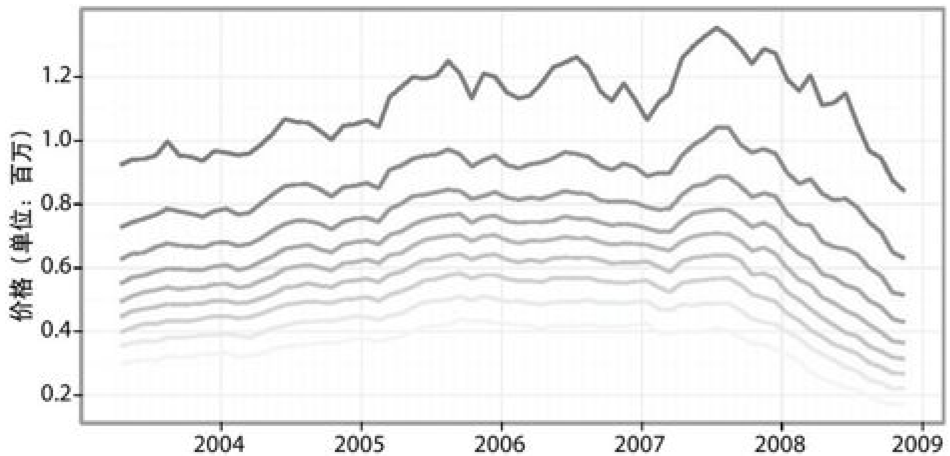
Joaquin Santa Cruz 2008

[illegible]

[illegible]“[illegible]”(huses)[illegible]

[illegible]10%[illegible]20%[illegible]30%[illegible]40%[illegible]50%[illegible]60%[illegible]70%[illegible]80%[illegible]90%[illegible]9[illegible]

[illegible]18-3[illegible]90%[illegible]10%[illegible]2007[illegible]



[illegible] 18-3[illegible]

[illegible]

[illegible]18-4[illegible]1.0[illegible]

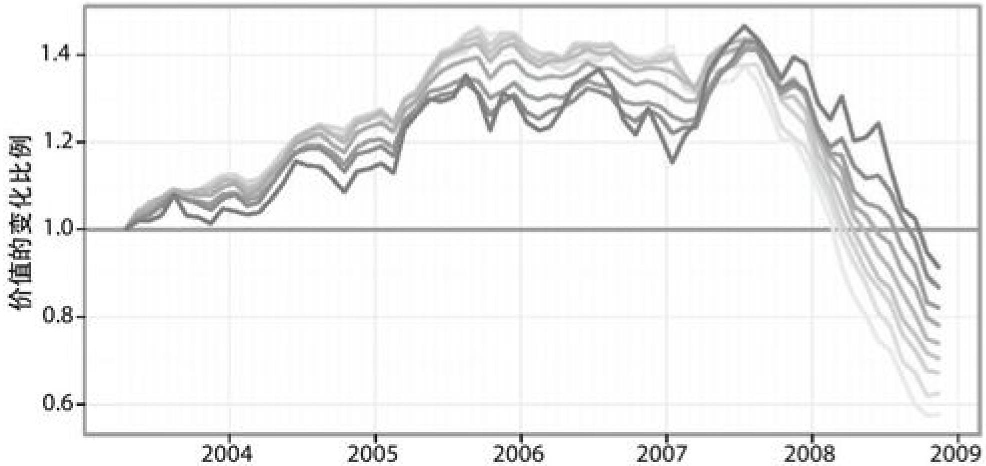
[illegible]2005[illegible]2007[illegible]

[illegible]2003[illegible]43%[illegible]9%[illegible]18-3[illegible]18-4[illegible]

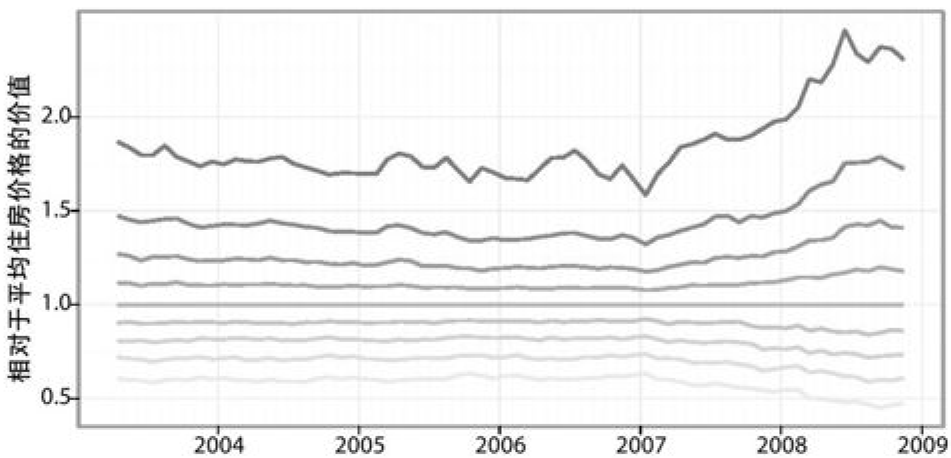
[illegible]18-5[illegible]

[illegible]1.2[illegible]20%[illegible]0.8[illegible]20%[illegible]2007[illegible]

[illegible]


价值的变化比例
1.4
1.2
1.0
0.8
0.6
2004
2005
2006
2007
2008
2009

[illegible] 18-4[illegible]

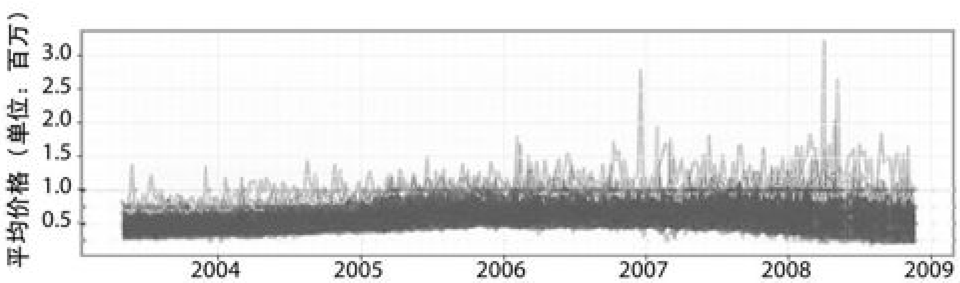

相对于平均住房价格的价值
2.0
1.5
1.0
0.5
2004
2005
2006
2007
2008
2009

[illegible] 18-5[illegible]2007[illegible]

[illegible]

[illegible]10[illegible]58[illegible]245[illegible]24%[illegible]428415[illegible]82%[illegible]

[illegible]18-6[illegible]“[illegible]”[illegible]



[illegible] 18-6[illegible]“[illegible]”[illegible]“[illegible]”[illegible]

[illegible](GM)[illegible](Wod 2006[illegible]

[illegible]18-7[illegible]

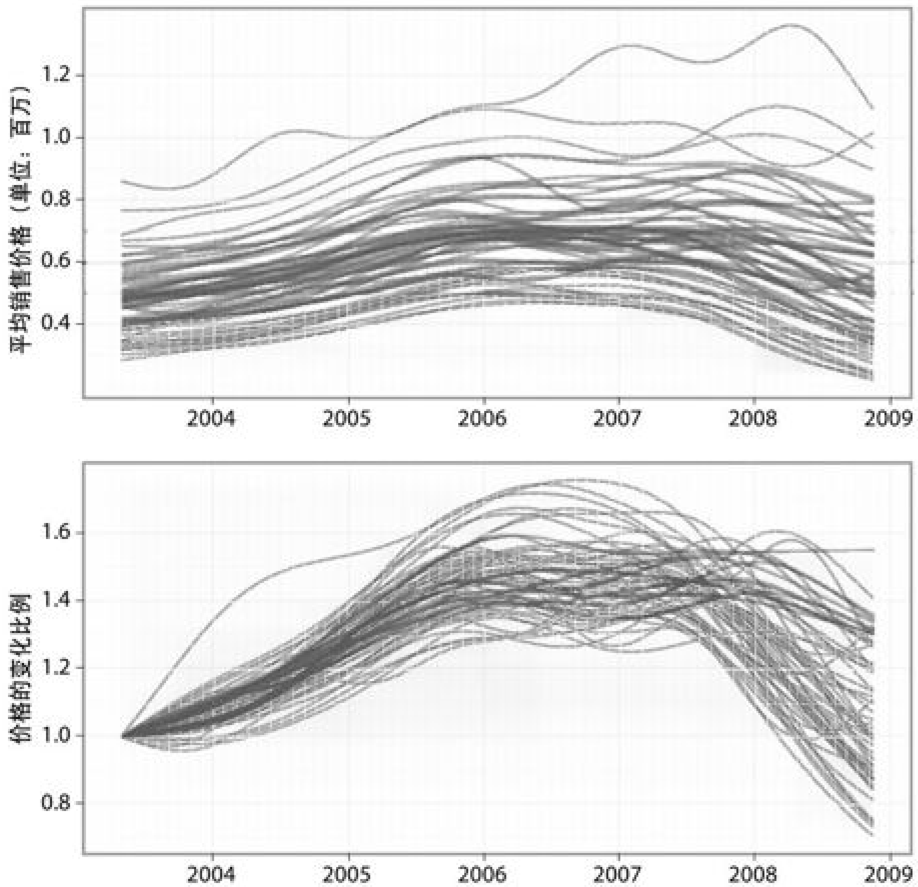
[illegible]

[illegible]18-7[illegible]

[illegible]2007[illegible]18-8[illegible]Mountain View[illegible]Oakley[illegible]Vallejo[illegible]San Pablo[illegible]

[illegible]San Jose[illegible]San Jose[illegible]Berryessa[illegible]San Jose[illegible]San Jose[illegible]San Jose[illegible]Cambrian[illegible]San Jose[illegible]San Jose[illegible]

[illegible]“[illegible]”[illegible]2006[illegible]2[illegible]2008[illegible]11[illegible]“[illegible]”[illegible]18-9[illegible]



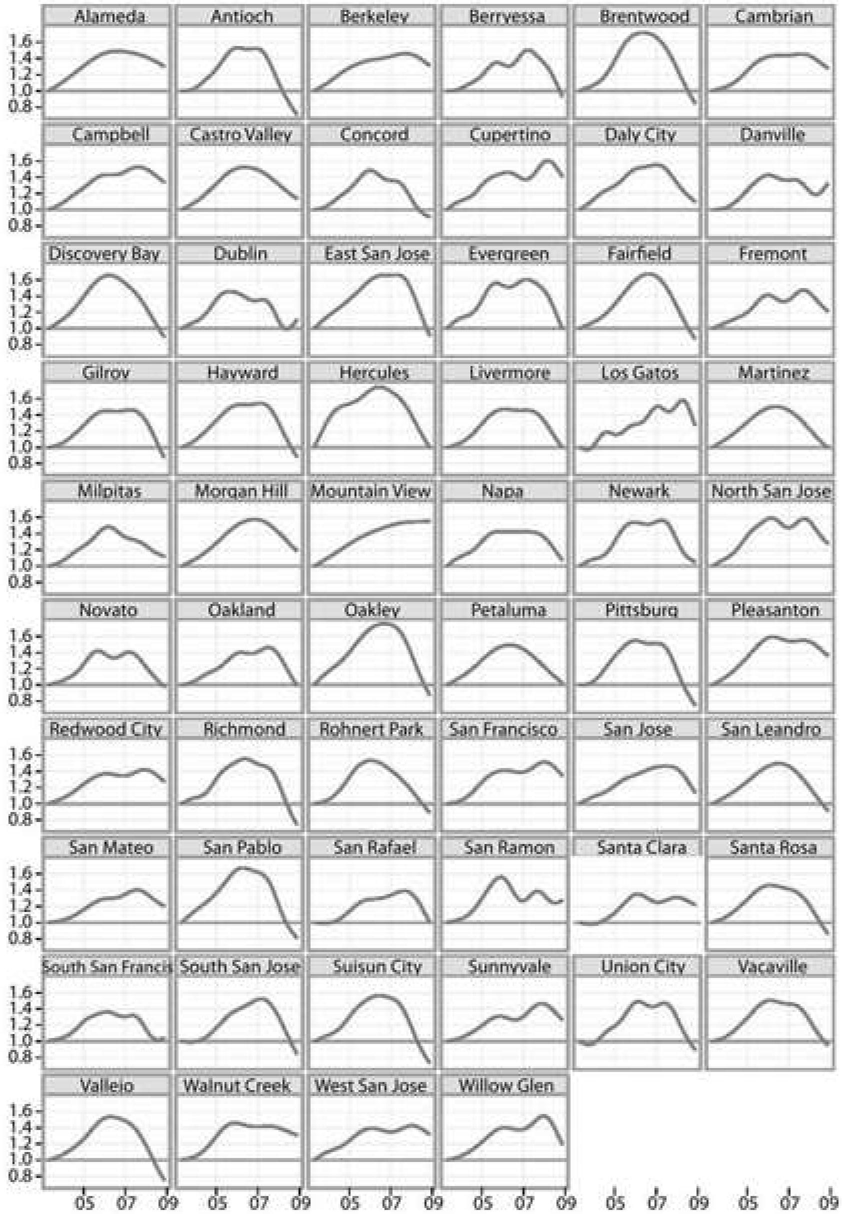
平均销售价格（单位：百万）
1.2
1.0
0.8
0.6
0.4
2004
2005
2006
2007
2008
2009
价格的变化比例
1.6
1.4
1.2
1.0
0.8
2004
2005
2006
2007
2008
2009

18-7

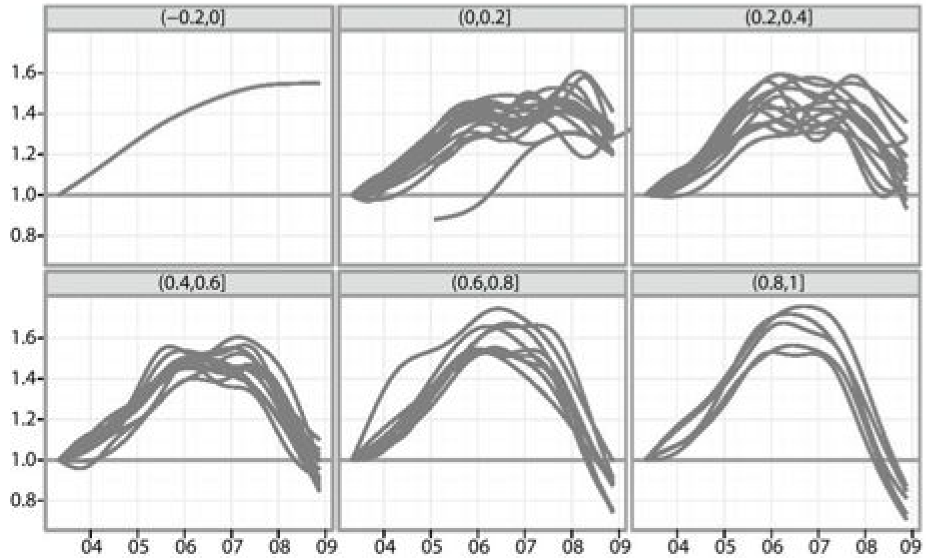
18-10

[illegible]

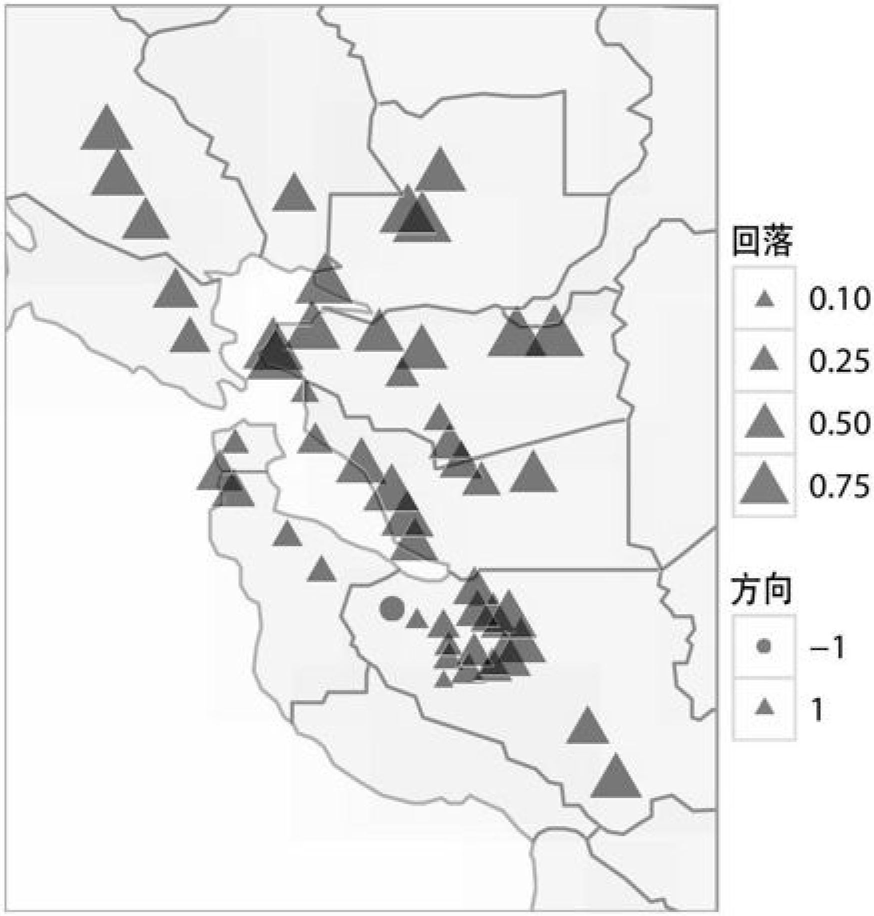
[illegible]quickfacts[illegible]
http[illegible]//quickfacts.census.gov/qfd/states/06/0649670.html)[illegible]
[illegible]csv[illegible]58[illegible]46[illegible]San Jose[illegible]“[illegible]”[illegible]



[illegible] 18-8 [illegible] 18-7 [illegible] San Pablo [illegible] Mountain View [illegible]



[illegible] 18-9[illegible]0.8[illegible]1[illegible]80%[illegible]100%[illegible](Muntain View)[illegible]



[illegible] 18-10[illegible]Mountain View[illegible]

[illegible]18-11[illegible]35[illegible]

[illegible]18-12[illegible]2000[illegible]2006

□□□2008□□□

□□

□San Joaquin□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

□□□□□□□□□□□□□□□□□□□San Benito□Santa Cruz)□□□□□□□□□□□

□□□□□□□□□□18-12□□□□□□□□□□□□□□□□□□□□□□□□□□2000□

2006□□□□□□□□□□□□□□2005□□□□□□□□□□□□□□□□□□□□□□□□□□

□□

□□□□□□□□□□□□□□□

□□□□□□□□□□□□□□□18-11□□□□□□□□□□□

□□□New York Times□□□□□□(MKinley 2007□□Stockton□□□

San Joaquin□□□□□□□□□□2007□□□□□□□□□□□□□□□□□□□□□□□□□

□□□□□□□□□Stockton□□□2008□□□□□□□□□□□□2008□□□□□□□

□□□□□□□□□□□□□□Stockton□□□□□□□□10□□□□□□□□□□□□□□□□□

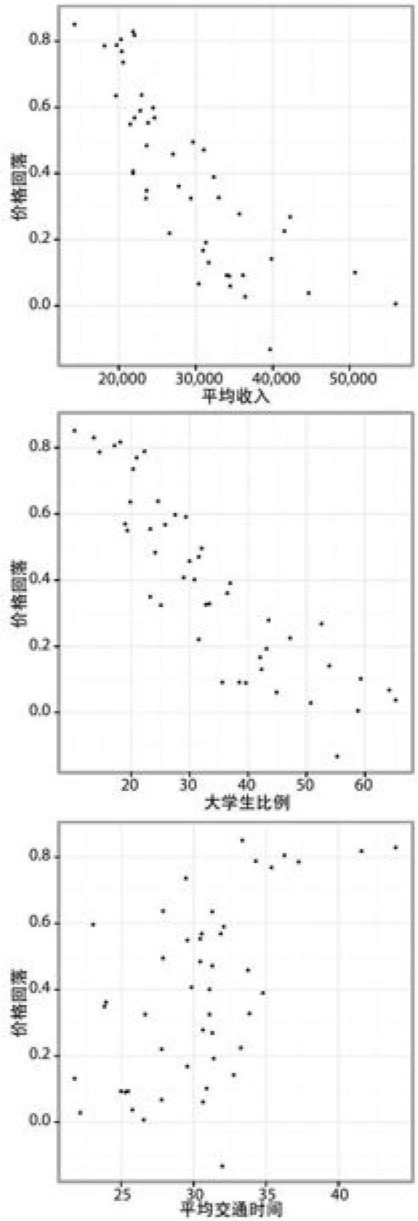
□□

□□□□□□□□□□□□Modesto□Merced□□□□□□□□(Cntral Valley)□□□

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

[illegible]

[illegible]25377[illegible]18-13[illegible]



0.8
0.6
0.4
0.2
0.0
价格回落
20,000
30,000
40,000
50,000
平均收入
0.8
0.6
0.4
0.2
0.0
价格回落
20
30
40
50
60
大学生比例
0.8
0.6
0.4
0.2
0.0
价格回落
25
30
35
40
平均交通时间

[illegible] 18-11[illegible]

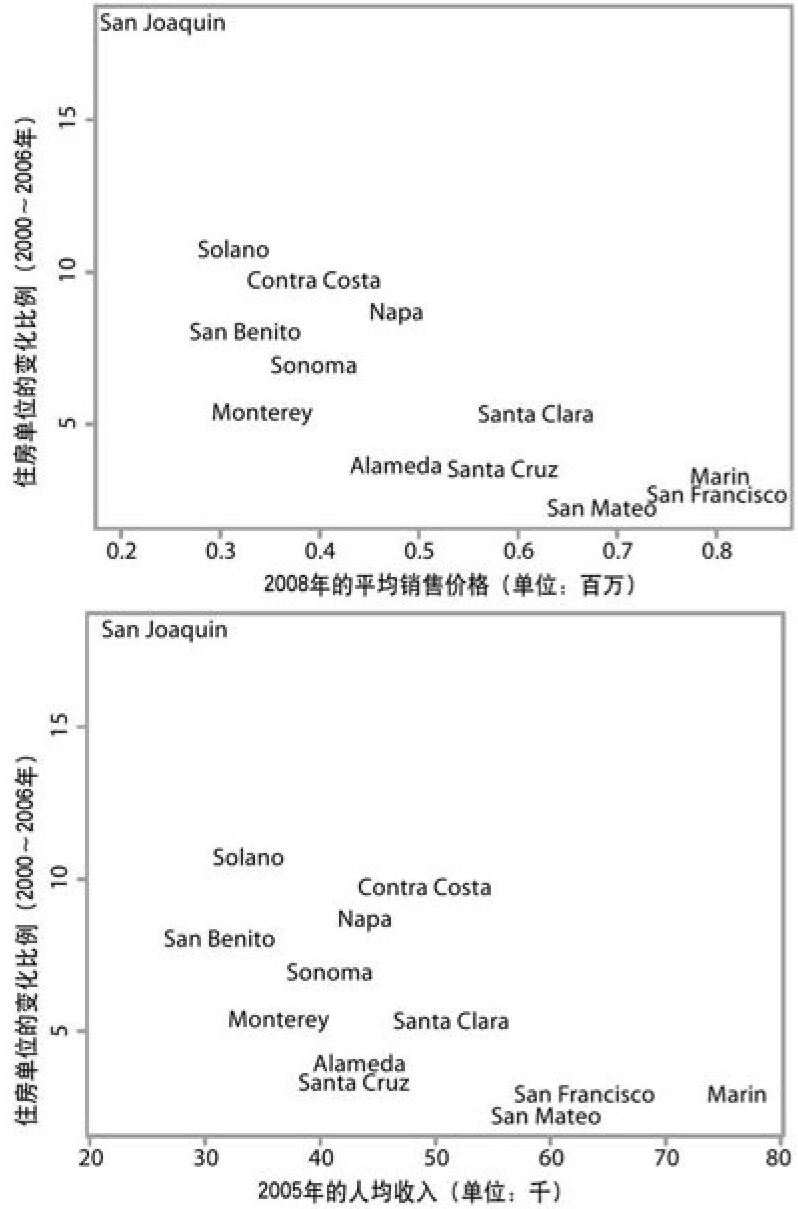

San Joaquin
Solano
Contra Costa
Napa
San Benito
Sonoma
Monterey
Santa Clara
Alameda
Santa Cruz
Marin
San Francisco
San Mateo
住房单位的变化比例（2000～2006年）
2008年的平均销售价格（单位：百万）
San Joaquin
Solano
Contra Costa
Napa
San Benito
Sonoma
Monterey
Santa Clara
Alameda
Santa Cruz
San Francisco
Marin
San Mateo
住房单位的变化比例（2000～2006年）
2005年的人均收入（单位：千）

图 18-12 [illegible]

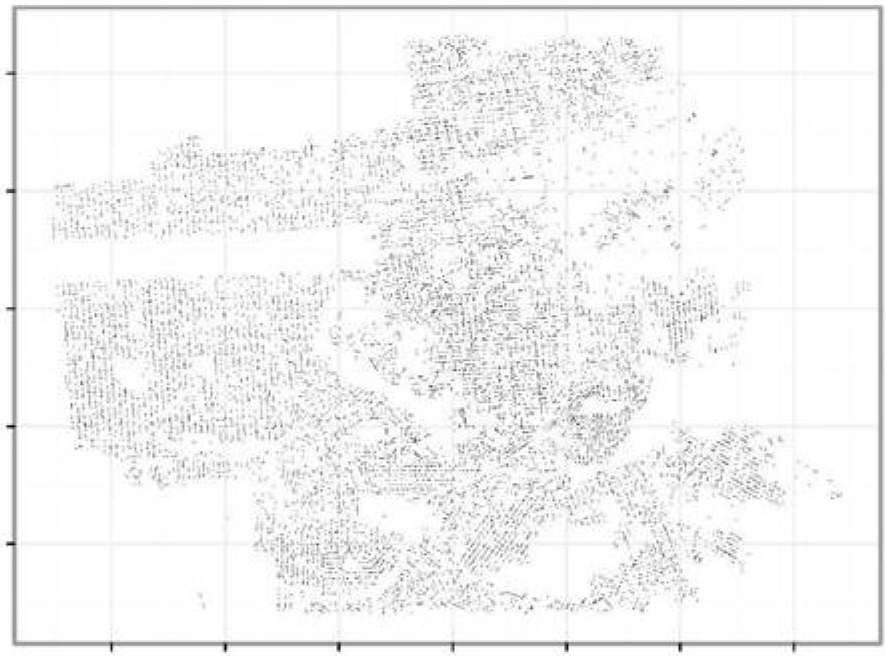
[illegible]

[illegible]18-14[illegible]

[illegible]

[illegible]

0.005[illegible]

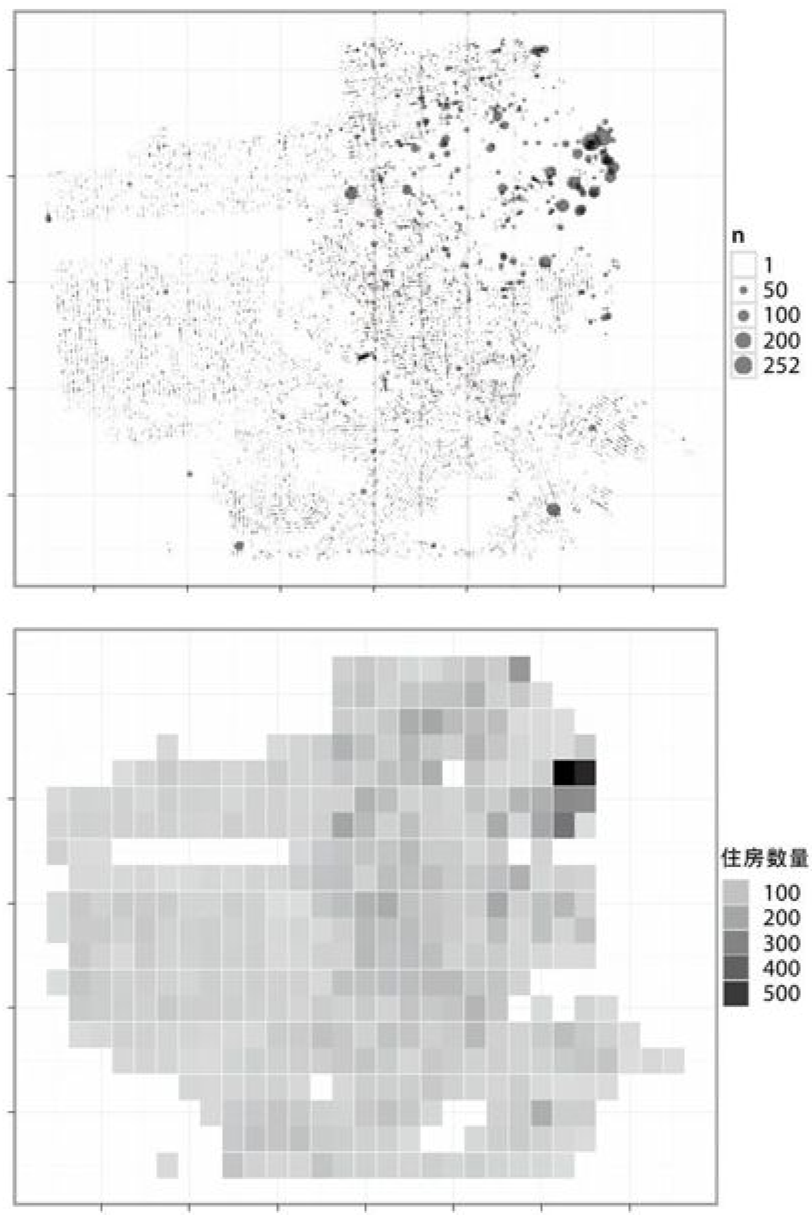




□　18-13□□

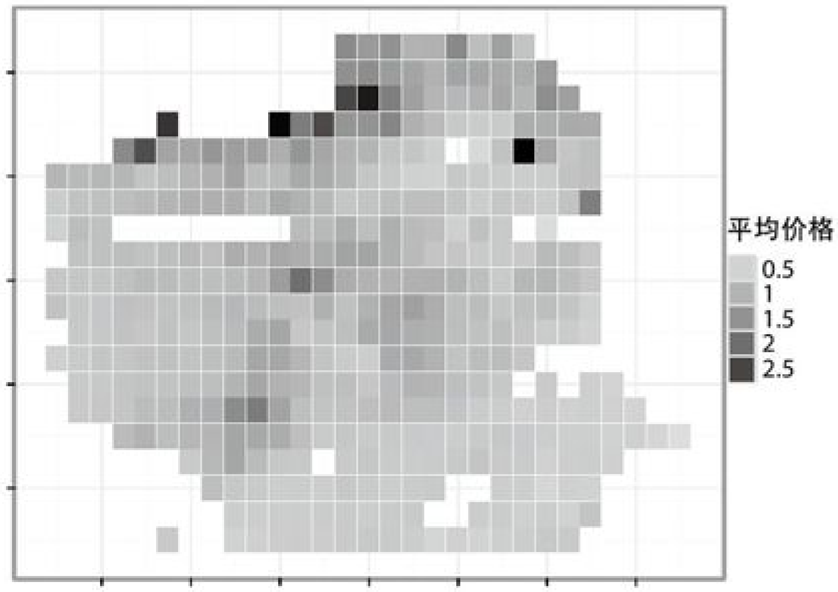
□□□□□□□□□□□□□□□□□□□□□□□□□□□□□
http□//openstreetmap.com□□□□68□

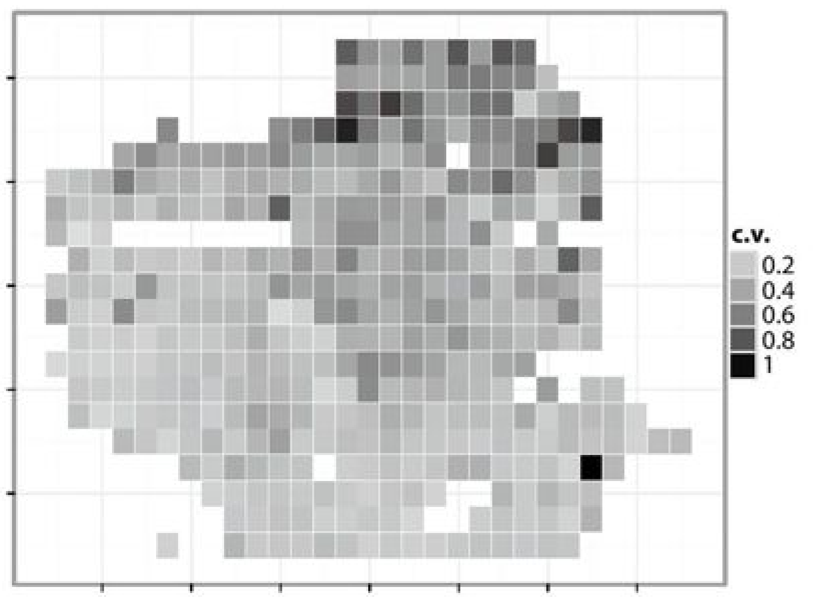
□□$100000□□□□□□□□□□□□□□□□□□□□□□□□□□□□



[illegible] 18-14[illegible]

[illegible]18-15[illegible]Presidio[illegible]St.Francis Wood[illegible]





[illegible] 18-15[illegible]

[illegible]

[illegible]

[illegible]

[illegible]2008[illegible]

[illegible]git[illegible]
https：//github.com/hadley/sfhousing[illegible]
[illegible]R[illegible]Perl[illegible]shell[illegible]
[illegible](Gntleman[illegible]Temple Lang 2007[illegible]

[illegible]

参考文献

[1]Gentleman,Robert and Duncan Temple Lang. 2007."Statistical analyses and reproducible research."Journal of Computational and Graphical Statistics，16（1）：1-23.

[2]Goldberg,Daniel W. 2008."A geocoding best practices guide."Technical report,GIS Research Laboratory,University of Southern California.http：//www.naaccr.org/filesystem/pdf/Geocoding_Best_Practices.pdf.

[3]Goldberg,D. W.，and J.P.Wilson.2008.USC WebGIS Services.https：//webgis.usc.edu.Last accessed December 2008.

[4]McKinley,Jesse."From housing to haven to foreclosure leader."New York Times. August 13，2007.

[5]Wood,Simon. 2006.Generalized Additive Models:An Introduction with R.Boca Raton,FL:Chapman＆Hall/CRC.

[illegible]19[illegible] [illegible]

Andrew Gelman[illegible]Jonathan P. Kastellec[illegible]Yair Ghitza

[illegible]“[illegible]”(satistics)[illegible]“[illegible]”[illegible]“[illegible]”[illegible]Playfair[illegible]Laplace[illegible]Galton[illegible]20[illegible]Gallup[illegible]Svengali[illegible]1964[illegible]The 480[illegible]The Ugly American[illegible]Fail-Safe[illegible]USA Today[illegible]New York Times[illegible]FiveThirtyEight.com[illegible]

[illegible]

[illegible]

·“[illegible]X[illegible]……”

·“[illegible]……”

·“[illegible]……”

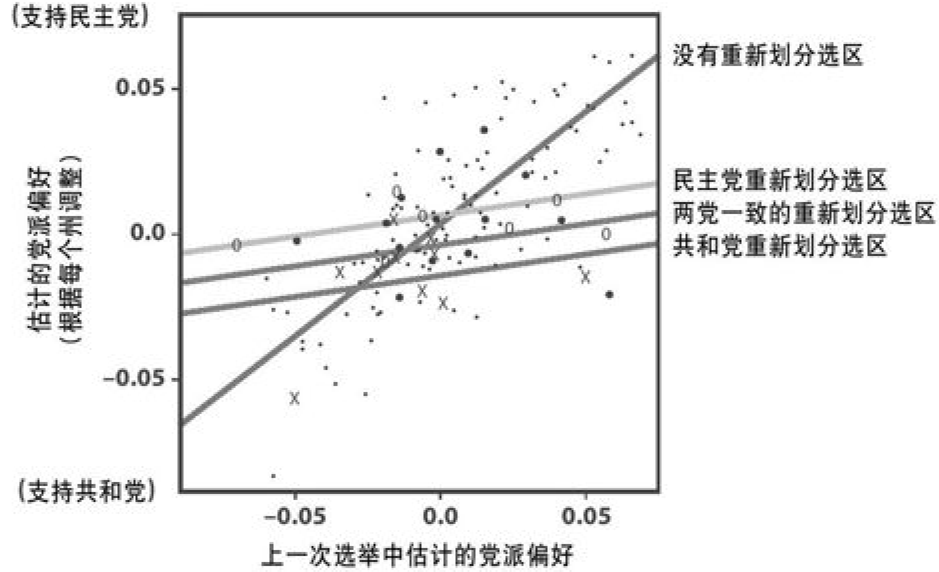
·“[illegible]……[illegible]……”

·“[illegible]……”

[illegible]——[illegible](Glman 2003[illegible]——[illegible]

[illegible]1[illegible]

[illegible]19-1[illegible]1972[illegible]Missouri)[illegible]



[illegible] 19-1[illegible]“[illegible]”[illegible]“[illegible]”[illegible]0[illegible]

“[illegible]”[illegible]50%[illegible]-5%[illegible]5%[illegible]45%[illegible]55%[illegible]

“ ” “ ” 10

“ ” 60% 100% “ ”

“ ”

19-1 (Glman King 1994

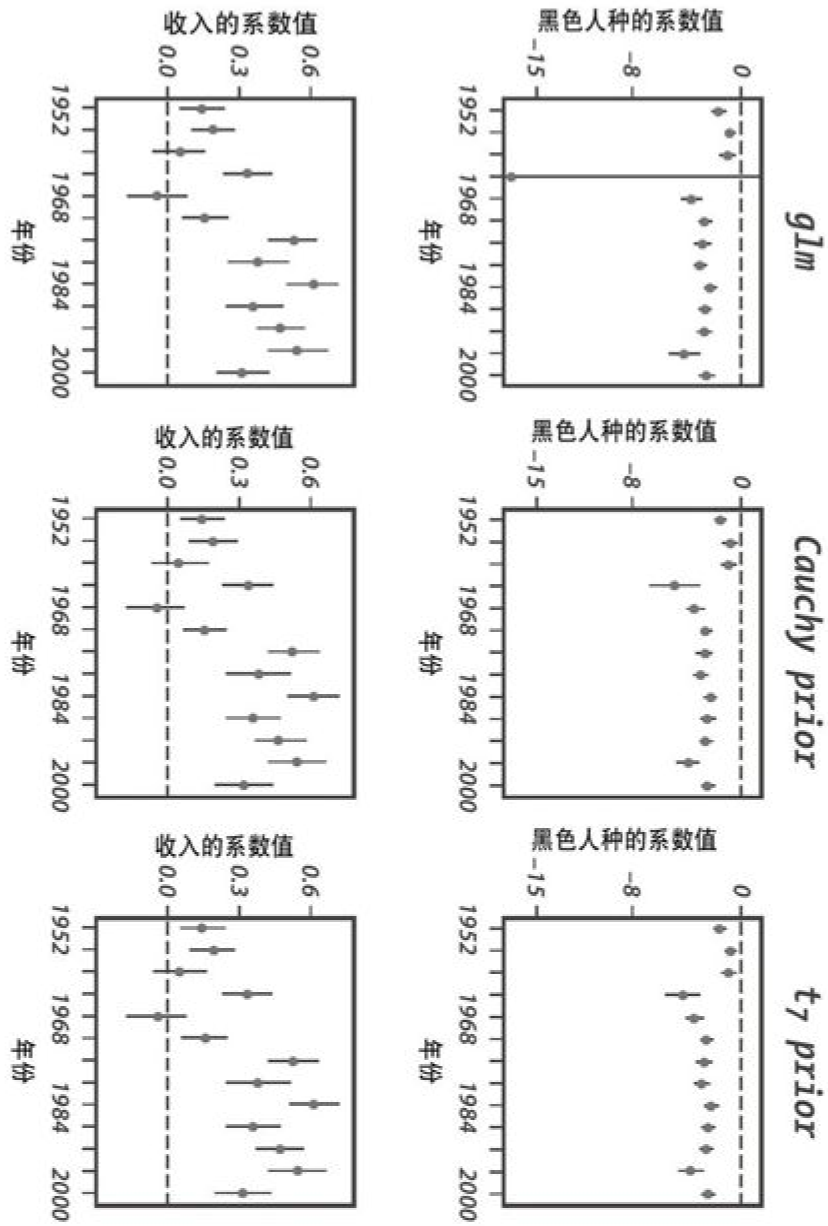
“ ”

[illegible]2[illegible]

[illegible]19-2[illegible]“[illegible]/[illegible]”[illegible]1952[illegible]2000[illegible]

[illegible]

[illegible]19-2[illegible]1964[illegible]——[illegible]0%[illegible]1964[illegible]Barry Goldwater[illegible]the Civil Rights Act[illegible]0[illegible]19-2[illegible]



□ 19-2□□□□□□□□□□□□□□□□□□□□□□□□□±1□□□□□□□□□□□□□□□□□□□□□□□□□□□□□1952□2000□□□□□□□□□□□□□□□□□□□□□□□1-5□□□□□□□□□□□□□□□□□□□□□□1964□□□□□□□□□□□□□□□□□□□□□□□Barry Goldwater)□□□□□□□□□□□□-1□□□□□□□□□□□□□□□□□□□□□R□□glm□□

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□“□□□”□□□□□□□□□□

□□□□□□□□1964□□□□□□□□□□□□□□□□□□□-1□□□□□□□□□□□□□□□

□□□□

□□□□□□□□□□□□□□□□□□□□□□□□□1990□Tufte□1967□Bertin□

□□“□□□”□□□□□□□——□□□□□□□□——□□□□□□□□□□

□□□□□□□□□19-2□□□□□□□□□□1964□□□□□□□□□□□□□□□□——□

□1952□2000□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

□□

□□

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□(Glman□

2002□□

□□

□□

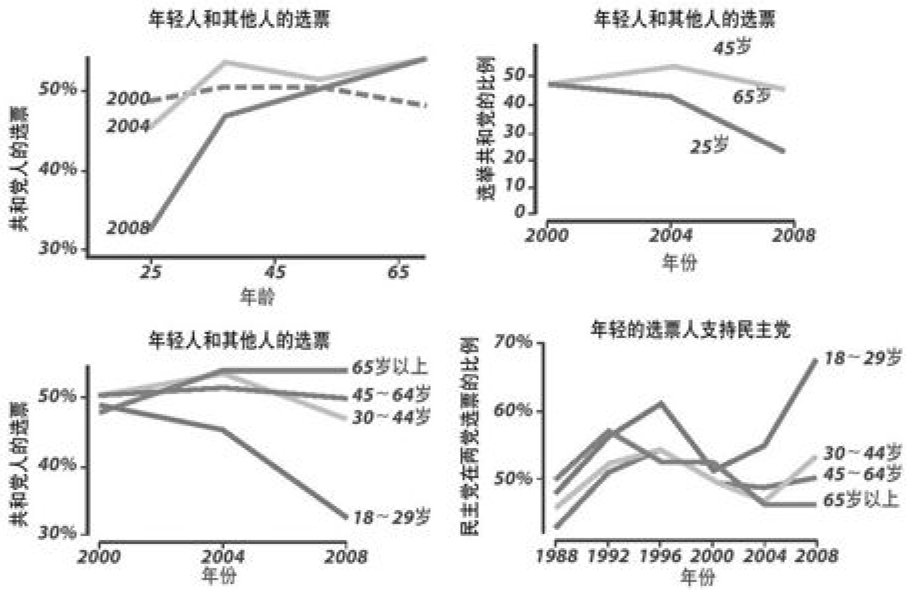
□□□□□□□□

[illegible]3[illegible]

[illegible]Mark Penn[illegible]New York Times[illegible]“[illegible]John Kerry[illegible]”Penn[illegible]

[illegible]19-3[illegible]CNN[illegible]Web[illegible]2008[illegible]Mark Penn[illegible]——[illegible]

[illegible]



□ 19-3□□□□□□□□□□□□□□□□□□□□□□□□□□□□

□□Hober Short□□□□Hober Short□□□□□□□□Web□□□□□□□□□□□□□□□□x□□□□□□□□□□□□□Short□□□“□□□”□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□2008□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□1988□□□□□1996□Bill Clinton□□□□□□□□——□□□□□□□□□□□□□□□□□□□□□□□□□□□□□——□□□□□□□□□□2008□□□□□□□□

□□□——□□□□□□□□□2008□□□

□□

[illegible]4[illegible]

[illegible]

[illegible]

[illegible]

(Kstellec[illegible]2008[illegible]

[illegible]19-4[illegible]

[illegible]

[illegible]

[illegible]“[illegible]”[illegible]“1”[illegible]“0”[illegible]

[illegible]

[illegible]

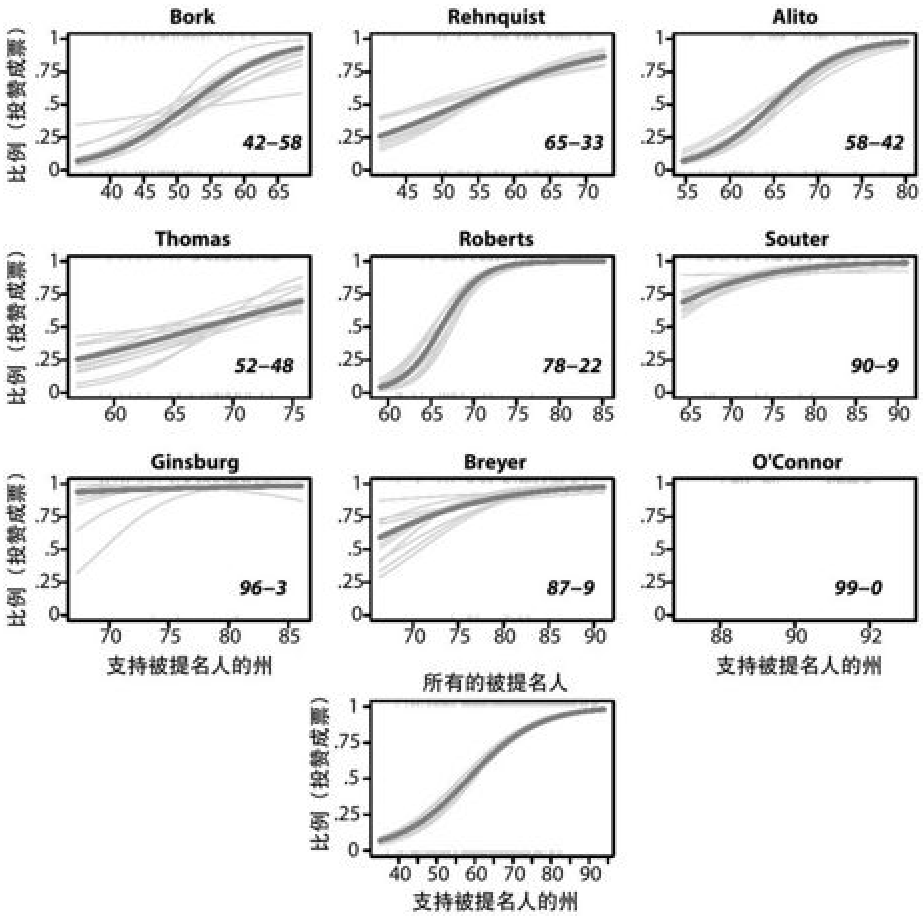

Bork
Rehnquist
Alito
Thomas
Roberts
Souter
Ginsburg
Breyer
O'Connor
所有的被提名人
比例（投赞成票）
支持被提名人的州
42–58
65–33
58–42
52–48
78–22
90–9
96–3
87–9
99–0

图 19-4[illegible]“1”[illegible]“0”[illegible]

[illegible]“[illegible]”[illegible]

[illegible]9[illegible]9[illegible]19-2[illegible]

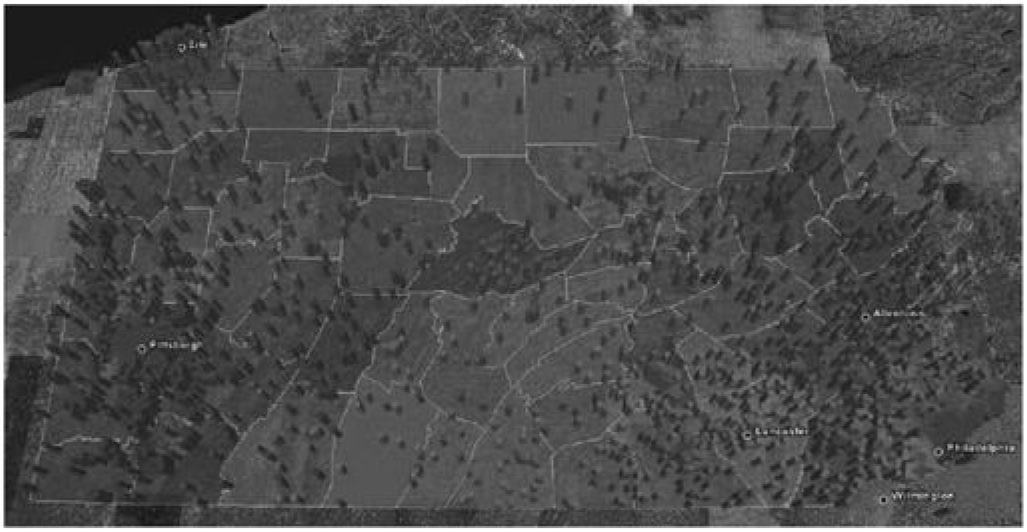
□□5□□□□□□□□□□□□

1986□□□□□□□□□□·□□□□□□□□□□□□□□□□□□□□□□□□□□□□Paoli□Penn Hills□□□□□□□Alabama□Paoli□□□□□□□□□Penn Hills□□□□□□□□□□□□□□□□□□□□□□“□□□”(sing state)□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□“□□”□“□□”□□□□□□□□□□□□21□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□2000□2004□□□□□□□□□□□□□□□□□□□□“□□”□□□□□□□□□□“□□”□□□□□□□□□□□□□□□□□□□□□□□□“□□”□□“□□”□□□□□□□□□□□□□□□□□□□□□□

□19-5□□2004□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□John Kerry□□□□□□□□□□□□□□□□□□George W.Bush□□□□□□□□□□□□□□□□□□□□□□□□□□□-□□-□□□

□□□□□□□——□□□□□□——□□□□□□□□□□□4000□□□1□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□1□□——□□□□□□□□□□□——□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□



[illegible] 19-5[illegible]2004[illegible]

[illegible]John Kerry[illegible]

George W.Bush[illegible]

4000[illegible]1[illegible]

[illegible]1[illegible]

[illegible]——[illegible]——[illegible]

[illegible]

69[illegible]

[illegible]

[illegible]Catalist[illegible]Catalist[illegible]

[illegible]

[illegible]Web[illegible]NationMaster.com[illegible]Name Voyager(http[illegible]//www.babynamewizard.com/voyager)[illegible]Hans Rosling[illegible]TED[illegible]

□□□□

[1]Bertin,J.□1967□. Semiology of Graphics.Translated by W.J.Berg□1983□.Madison:University of Wisconsin Press.

[2]Gelman,A.□2003□."A Bayesian formulation of exploratory data analysis and goodnessof-fit testing."International Statistical Review 71□369-382.

[3]Gelman,A.□A. Jakulin,M.G.Pittau,and Y.S.Su□2008□."A weakly informative default prior distribution for logistic and other regression models."Annals of Applied Statistics,to appear.

[4]Gelman,A. and G.King□1994□."Enhancing democracy through legislative redistricting."American Political Science Review 88□541-559.

[5]Gelman,A.□C. Pasarica,and R.Dodhia□2002□.“Let's practice what we preach:turning tables into graphs.”American Statistician 56□121-130.

[6]Kastellec,J.□J. Lax,and J.Phillips□2008□.“Public opinion and Senate confirmation of Supreme Court nominees.”Technical report,Department of Political Science,Columbia University.

[7]Tufte,E. R.（1990）.Envisioning Information.Cheshire,CT:Graphics Press.

□20□　□□□□

Toby Segaran

□□Andreas Weigend□□“□□□□‘□□□□□□□□□□□□□□□□□’□□□□□□□□□‘□□□□□□□□□□□□’”□

□□□□□□□□□□□□□□TB□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□“□□”□□□——□□□10□□

□□□□□“□□□”□□□——□□□□□□□□□□□□□□□

□□□

[illegible]

[illegible]Freebase(http[illegible]//www.freebase.com)[illegible]

[illegible]

[illegible](htp[illegible]//opensecrets.org)[illegible]

·[illegible]http[illegible]//www.census.gov[illegible]

·Geonames(http[illegible]//www. geonames.org)[illegible]

·[illegible](htp[illegible]//sec.gov)[illegible]

·[illegible](htp[illegible]//epa.gov)[illegible]

·[illegible](htp[illegible]//uspto.gov)[illegible]

·[illegible]

·[illegible](htp[illegible]//usda.gov)[illegible]

·[illegible](NBI[illegible]http[illegible]//ncbi.nlm.nih.gov)[illegible]Genbank[illegible]Pubmed[illegible]dbSNP[illegible]

·[illegible]

·[illegible](htp[illegible]//medicare.gov)[illegible]

(htp[illegible]//www.fda.gov)[illegible]

·[illegible]——[illegible]

[illegible]

[illegible]20[illegible]80[illegible]Wall Street[illegible]Charlie Sheen[illegible]Michael Douglas[illegible]Douglas[illegible]Sheen[illegible]Sheen[illegible]Douglas[illegible]

[illegible](Scurities and Exchange Commission,SEC)[illegible]

[illegible]

[illegible](Uited States Department of Agriculture,USDA)[illegible](EA)[illegible]

[illegible]Geonames[illegible]

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□(Cnter for Responsible Politics,CRP)□□□20-1□□□□□□□□□□□□□□□□□□□□□□□□



□　20-1□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□70□

□□(Cntral Index Key,CIK)□□□□□□□Exxon Mobile□□□□□□□□□□□□□“□□”□□□□□□□□□“□□□□”□□

[illegible]

[illegible]Web[illegible]“[illegible]”[illegible]——[illegible]

[illegible]

[illegible]

[illegible]“[illegible]”[illegible]“[illegible]”[illegible]

[illegible]

[illegible]
[illegible]

[illegible]

[illegible]

[illegible]……

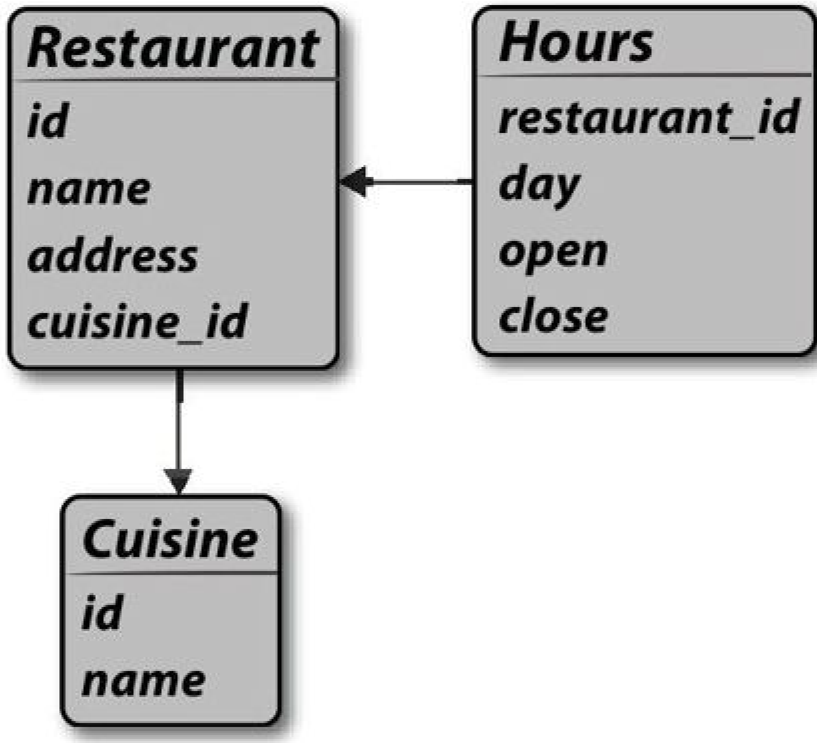
[illegible]

[illegible]Excel[illegible]

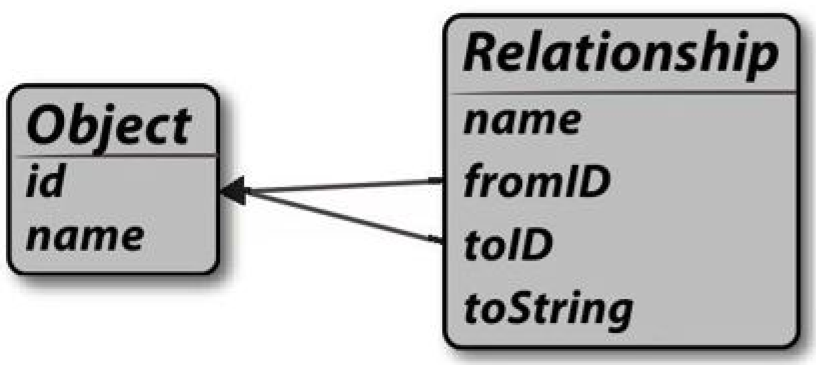
[illegible]20-2[illegible]

[illegible]

[illegible]20-3[illegible]UML[illegible]


Restaurant
id
name
address
cuisine_id
Hours
restaurant_id
day
open
close
Cuisine
id
name

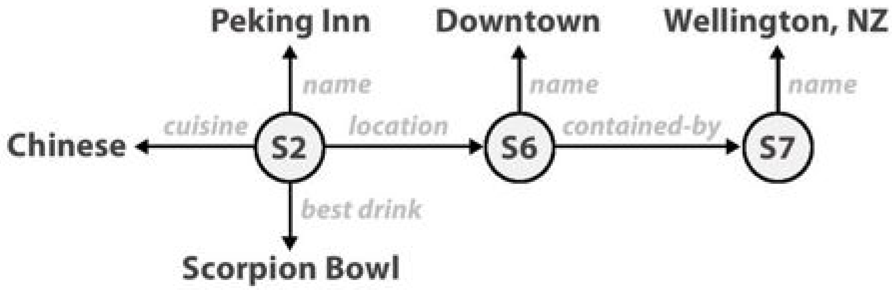
[illegible] 20-2[illegible]


Object
id
name
Relationship
name
fromID
toID
toString

[illegible] 20-3[illegible]

[illegible]

[illegible]20-4[illegible]



□ 20-4□□□□□“□□□□□”□□□□

□□

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□“□□□□□”□□□□□□□□□□□□□□□□□□

Sesame(http□//openrdf. org)

□□□□□□□□□□□□□□□□□Aduna□□□

Jena(http□//jena. sourceforge.net)

□□□□□□□□□□Hewlett Packard□□□

AllegroGraph(http□//agraph. franz.com)

□□□□□□□□□□□□□□□Franz□□□

Neo4J(http□//neo4j. org)

□□□□□□□□□□□□□□

□□□“□□□□□”□□□□□□□□□□□□□□□□□□□□□□□□□□□□□——□□□□□□□□

□□□□□□□

□□□“Coca-Cola”□□□□□□“The Coca-Cola Company”□□□□□□□“Coke”□□□□□□□“The Coca-Cola Bottling Company”□□□□□□□□□□□□□□□

□□□address□□□□□□□□□□□□□□□□location□□□□□□□□□□□□□□□□□location□□□□□□□□□□□□□□□□□□□□□□□□location□□□

□□□□□□□□□□□□□□□□□□□□□□□□□Web□□□□□“□□□□□□”□□□□□□□□□□□□□□□□□□□□(UI)□□□(otologies)□□□

□□□□□□□□□□□□□□□□URI□□URI□□

□□□□□□□□□

□□“Coca-

cola”[illegible]“The coca-cola company”[illegible]——[illegible]

[illegible]Netflix[illegible]Prêt-à-Porter[illegible]Netflix[illegible]Ready to Wear[illegible]B.U.S.T.E.D.[illegible]Everybody Loves Sunshine[illegible]Point of No Return[illegible]The Assassin[illegible]——[illegible]

[illegible]1994[illegible]Robert Altman[illegible]Julia Roberts[illegible]Sophia Lauren[illegible]——[illegible]——[illegible]

[illegible]Ghostbusters[illegible]Ghostbusters 2[illegible]1984[illegible]1989[illegible]——[illegible]The Madness of George III[illegible]The Madness of King George[illegible]

[illegible]

[illegible]“[illegible]”[illegible]

[illegible]

[illegible]7[illegible]“[illegible]”[illegible]City of God[illegible]50[illegible]“John Smith”[illegible]

[illegible]“Toby Segaran”[illegible]

[illegible]

[illegible]

[illegible]“[illegible]”[illegible]

[illegible]

[illegible]7[illegible]Jeff Jonas[illegible]Jeff[illegible]“Patrick Smith”[illegible]“Patricia Smith”[illegible]“Pat Smith”[illegible]

[illegible]“[illegible]”[illegible]“[illegible]”

[illegible]/[illegible]Prêt-à-Porter[illegible]Ready to Wear[illegible]“[illegible]”

[illegible]——[illegible]

□□□□

□□□□□□□□□□□□□□□□□□□□□□□“□□□□”(cllective reconciliation)□“□□□□□”(cllective entity resolution)□□□□□□□□□□□□Indrajit Bhattacharya□□□□□□□□http□//www.lib.umd.edu/drum/handle/1903/4241□□□□□

□□□

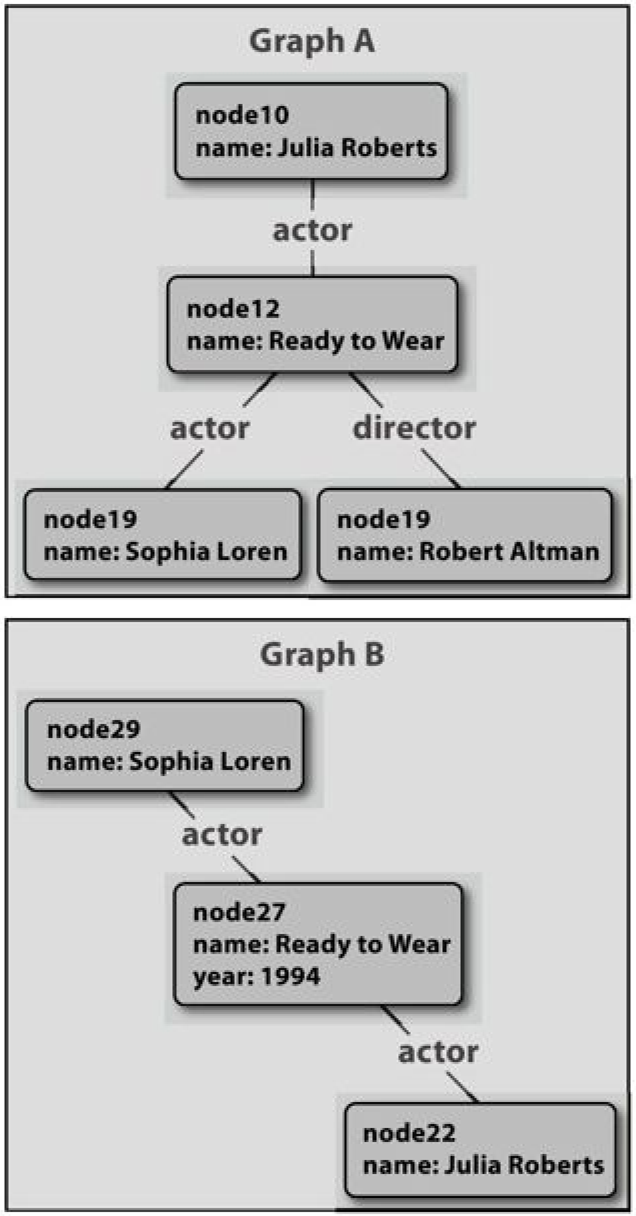
□□20-5□□□A□□□node10□B□□□node22□□□□□□□□□□□“Julia Roberts”□A□□□node12□B□□□node27□□□□□□□□□□□□“Ready to Wear”□

□□□□□□□□□□□□□□□□□□□□□□□□□□□□——□□□□“Julia Roberts”□□□□□□“Ready to Wear”□□Julia Roberts□□□□□□Ready to Wear□□□□□□□□□□□□□□□□□□□□□□□□□

□□□□□□□□□□□□□□□□□“□□□”(mssage-passing)□□□□A□□□node12□□□□□B□□□node27□□□□B□□□□□□□A□

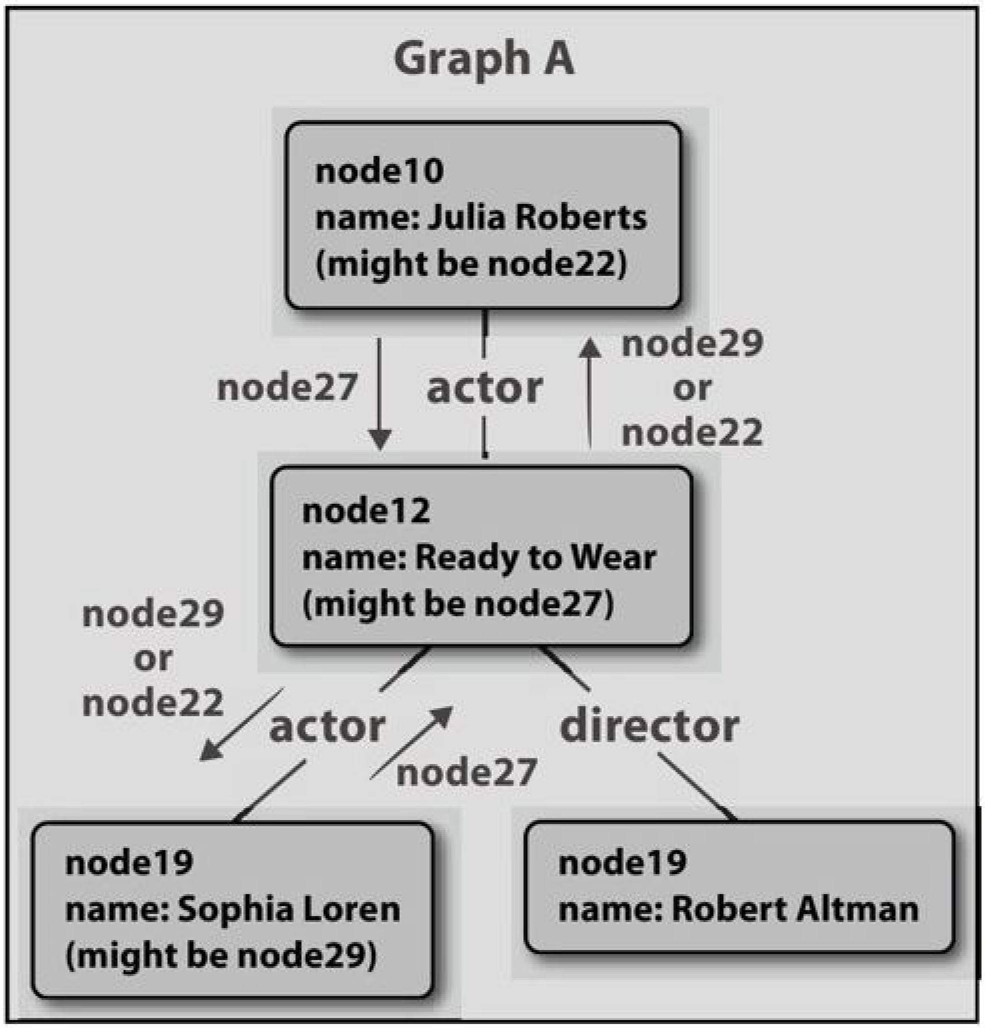
□□□□□□□□□□□□□□□□□□□□□□□“□□□□□□node22□□□node25□□”□□□□node10□□□□□□□□□□□□□“□□□□□□□□□□□□□□□node22□□”□□□□□□node10□□□□node12□□□□□□□□□□20-6□□□□□□□□□

□□□□□□□□□□□□□□□“□□□□”□□20-7□□□□□□□□□□



[illegible] 20-5[illegible]

[illegible]“Ready to Wear”[illegible]node10[illegible]Julia Roberts[illegible]Julia Roberts[illegible]“[illegible]”[illegible]“[illegible]”[illegible]



□ 20-6□□□□□□□□

1.□□node11□□□□□□node23□□□□□□□□□□□□□□□□

2.□□□□□□node15□□□□□□node9□□□□□□□□□□□□□□□Julia Roberts□□□□□□□□□□□□□□□□□□□□□

3.□□node11□□□□□□□□□□node10□“□□□□□□node22□□□□node24□□”□

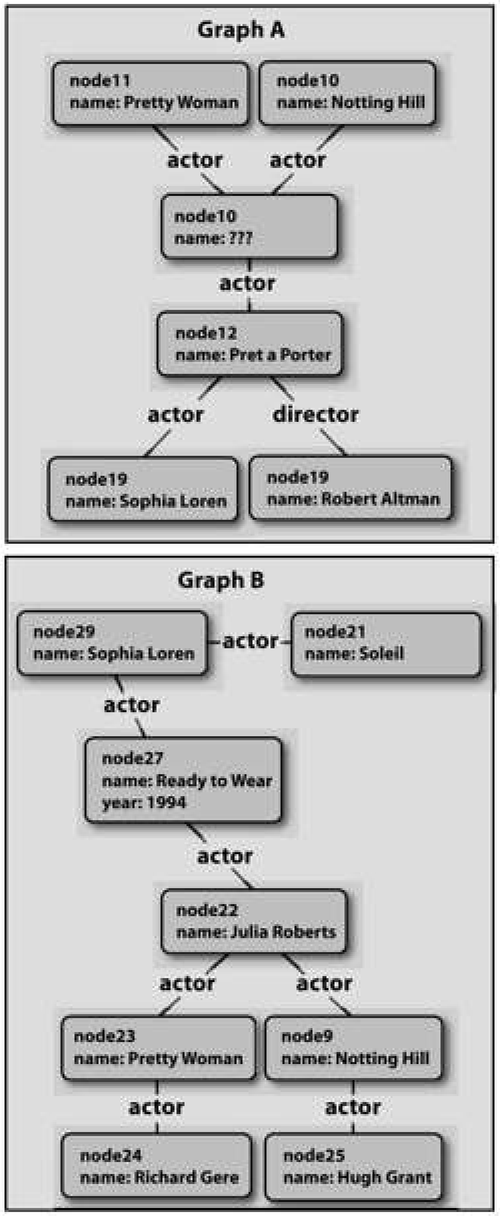
4.□□□□□node15□□□□□□□□□□node10□“□□□□□□node22□□□node25□□”□

5.□□node10□□□□□□□□□□□□□□□□□□□□□□□□node22□□□□□□□□□□□□□

6.□□□□□node10□□

7.□□node12□□□□□node10□□□node19□□□□“□□□□□□node23□□”□□□□□□□□□□□node23□□□

□□□□□□□□□□□□□□□□□□□□□□Prêt-à-Porter□□□Ready to Wear□□□□□□□□



Graph A
node11
name: Pretty Woman
node10
name: Notting Hill
actor
actor
node10
name: ???
actor
node12
name: Pret a Porter
actor
director
node19
name: Sophia Loren
node19
name: Robert Altman
Graph B
node29
name: Sophia Loren
actor
node21
name: Soleil
actor
node27
name: Ready to Wear
year: 1994
actor
node22
name: Julia Roberts
actor
actor
node23
name: Pretty Woman
node9
name: Notting Hill
actor
actor
node24
name: Richard Gere
node25
name: Hugh Grant

[illegible] 20-7[illegible]

[illegible]12[illegible]8[illegible]

[illegible]

[illegible]“[illegible]”[illegible]

[illegible]

□□ □□□□

Ben Blackburne□Wellcome Trust Sanger Institute□□□□□□□□□□□□□□□□□□

Jean-Claude Bradley□Drexel□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□UsefulChem□□□□□□□□2005□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□wiki□□□Web□□□□□□□□□□□□□□□□□□Jean-Claude□□□□“Open Notebook Science”□□□□□□□□“Open Science”□□□wiki□□□

Lukas Biewald□Dolores□□□□□□□□□CEO□□□□□□□□□□“□□□□”(cowdsourcing)□□□□□□□□Dolores□□□□□□□(htp□//blog.doloreslabs.com)□□□□□□□“□□□□”□□□□□□□□□□□□□□□□Dolores□□□□□□□□□Powerset□□□□□□□□□□□□□Powerset□□□□□□□□Yahoo!Japan□□http□//lukasbiewald.com/□□□□Lukas□□□□□□□□□□□□□□

Brian Cooper□□□□□□□□□□□□□□□□□□□□□□□□□□Georgia Tech□□

□□

□□P2P□□□□□□□□□□□□□□□□□□□□XML□□

Jason Dykes□20□□90□□□□□□□□□□□□□□□□□□□□□□□□□

□□□□□□□□□□□□□Tcl/Tk□SVG/JavaScript□Processing□□□□

□□□□□□□□□□□□□□□□□□□□□(htp□//gicentre.org)□□□□□

1990□□□□□□□□□□□□□□2000□□Leicester□□□□□□□Jason□

□□□□□□□□□□□□□(Iternational Cartographic Association

Commission)□□□□□

Jonathan Follett□Hot Knife Design□□□□□□□□□□□□□□□□

□□□□□□□□□□□□□□□□□□□□A List Apart□□Digital Web□□

□UXmatters□□□□□□□□Boston□□□□□Web□□□□□□□□□□□

□□□□□□□□□□□□□□□□□Jon□□□□□□□□□□□□□□□□□

Horizon□□□□□□□□□

Andrew Gelman□□□□□□□□□□□□□□□□□□□□□□Data

Analysis Using Regression and Multilevel/Hierarchical

Models□□□□□□□□□Red State,Blue State,Rich State,Poor

State:Why Americans Vote the Way They Do□□□□□□□□□□

□A Quantitative Tour of the Social Sciences□□□□□□□□□

Yair Ghitza□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

□□□□□□□Catalist□□Copernicus Analytics□□

Rajarshi Guha[illegible]NIH[illegible]

[illegible]Indiana University[illegible]

[illegible]QSAR[illegible]Web[illegible]

Alon Halevy[illegible]

[illegible]1999[illegible]Halevy[illegible]Nimble Technology[illegible]

[illegible]2004[illegible]Transformic[illegible]Deep Web[illegible]

[illegible]G[illegible]Halevy[illegible]Computing Machinery[illegible]

2000[illegible]Presidential Early Career

Award(PECASE)[illegible]Sloan[illegible]1999[illegible]2000[illegible]150[illegible]

[illegible]1993[illegible]

Jeff Hammerbacher[illegible]Cloudera[illegible]

Cloudera[illegible]Jeff[illegible]Accel Partners[illegible]Accel[illegible]

Facebook[illegible]Facebook[illegible]

[illegible]Hive[illegible]

Hadoop[illegible]Cassandra[illegible]P2P[illegible]

Facebook[illegible]Jeff[illegible]Jeff[illegible]

Jeffrey Heer[illegible]

[illegible]

[illegible]prefuse[illegible]flare[illegible]

[illegible]Xerox PARC[illegible]IBM Research[illegible]

Microsoft Research□Tableau Software□□□□□□□□□□□□□□□□□□□
□□□□□□□□□□□□

Matthew Holm□□□□□Hot Knife Design□□□□□□□□□□□□□□□□□
□□□□□HTML/CSS□□□□□CMS□□□Web□□□Matt□□□Oregon□□□□□□
□□□□□(te Oregon chapter of the Association of Computing
Machinery's Special Interest Group on Computer-Human
Interaction,CHIFOO)□□□□□□□□□□□□□Matt□□□□□□□□□□□□□□□□□
□□□□□□□□Babymouse□(Rndom House□□□□□□□□□□□□□□□□□□□
□□□100□□□

J. M.Hughes□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□
2003□2007□□
□□□□□□□□□□□□□□□□□□□□NASA□□□□□□□□□□□□□□□□□□□□□□□□□□
□□□Tucson□

Jeff Jonas□IBM□□□□□□IBM□□□□□(IM Entity Analytics Group)
□□□□□□□IBM□□□□□□□□□□□□□□□□□□□□□□□Jonas□1984□□□□
2005□□IBM□□□Jonas□□□□http□//jeffjonas.typepad.com□

Jonathan P. Kastellec□□□□□□□□□□□□□□□□□Journal of Law□□
□Economics□Organization□□□Journal of Empirical Legal
Studies□□□Perspective on Politics□□□□□□□

Valdean Klump□□□□□□□□□□□□□□□□□□□□□□□□□□

Aaron Koblin□□□□□□□□□□□□□□□□□□□□Flight Patterns□The Sheep Market□Ten Thousand Cents□□□□□□□□□“House of Cards”□□□□□□□□□□□□□□□□□□□□□□□

Coco Krumme□MIT Media Lab□□□□□□□□□□□□□Metaweb Technologies□□□□□

Andrew Lang□Oral Roberts□□□

Pierre Lindenbaum 2000□□□□□□□□□□□□□□□□virushost□□□□□□□□□□□□□□□□French National Center of Genotyping□□□□□□□□2001□□□□French startup Integragen□□□□□□□□□□□□□□□□□□□□□□Fondation Jean Dausset-CEPH□□□

Jayant Madhavan□□□□□□□□□□□Deep Web□□□□□□□□□□□□□□□□□□Tranformic□□□2005□□G□□□□□□□□□□□Tranformic□□□Deep Web□□□□□□□Madhavan□□2005□□□□□□□□□□□□□□□□□□

Michal Migurski□Stamen Design□□□□□□□□□□□□□□□□□□□□□□□□□1995□□□□□□□Web□□□Web□□□http□//mike.teczno.com□□□□□□□□□□□□□□□□

Cameron Neylon[illegible]Science and Technology Facilities Council(STFC)[illegible]ISIS Neutron Scattering facility[illegible]Web[illegible]

Peter Norvig[illegible]AAAI[illegible]ACM[illegible]Artificial Intelligence:A Modern Approach[illegible](Pentice Hall[illegible]NASA[illegible]USC[illegible]Berkeley[illegible]

Brendan O'Connor[illegible]Dolores[illegible]Powerset[illegible]“Artificial Intelligence and Social Science”[illegible]http[illegible]//anyall.org/blog[illegible]

David Poole[illegible]AT[illegible]T[illegible](Aerican Statistical Association)[illegible]

Raghu Ramakrishnan[illegible]Audience and Cloud Computing[illegible]——[illegible]Web[illegible]——[illegible]SQL[illegible]1999[illegible]Birch[illegible]SIGMOD 10-Year Test-of-Time[illegible]Database Management Systems[illegible]Johannes Gehrke[illegible]McGraw-Hill[illegible]ACM SIGMOD[illegible]ACM[illegible]IEEE[illegible]

Toby Segaran□O'Reilly□□□□□□□□□□Programming Collective Intelligence□□□□□□□Programming the Semantic Web□□□□□□□□Metaweb□□□Metaweb□□□□□□□□□□□□□□□2003□□Genstruct□□□Genstruct□□□□□□□Toby□MIT□□□□□□□□□□□□□□□Brooke□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□http□//blog.kiwitobes.com□

Lisa Sokol□□□IBM□□“□□□□□□□”(Ationable Information)□□□Sokol□□□□□□□□□□□□□□□□□Massachusetts□□□□□□□□□□□□□

Utkarsh Srivastava□□PNUTS□□□□□□Pig□□□□□□□□□□Pig□□Hadoop□□

Deborah Swayne□AT□T□□□□□□□□□□□□□□□□□□□□□□□□□□□ASA□□□□□□□□□□□□□□□□□□□□□□□□ggobi□□□□□

Jud Valeski□Gnip□□□CTO□□□□□□Gnip□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□Jud□□20□□□□□□□□□□IBM□

Netscape□onebox.com□AOL□me.dium□□□□□□□□□□M□A□□□□□□□□□□□□□□□□□□□□□Jud□□□□□□□□□

Hadley Wickham□Rice□□□15□□□□□R□□□□□□2006□□□□□ggplot□□□□□□□□□□□□□□□R□□□□□Hadley□□□□□□□□John Chambers Award□

Antony Williams□ChemZoo□□□□□□□□ChemSpider□□□Advanced Chemistry Development(ACD/Labs)□□10□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□NMR□□□□□□□□100□□□□□□□□□Eastman Kodak Company□NMR□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□ChemSpider□□□□□□□□□□□□□□□□□□□□□

Egon Willighagen□□□Uppsala□□Chemistry Development Kit□Metware□□□□□□□□□□□□□□□□□□□□□10□□□□□□□□□□Jmol□Bioclipse□

Jo Wood□□□□□□□giCentre(http□//gicentre.org)□□□□□□□□□□□1990□□□□□□□□□□□□□□□□GIS□□□□□LandSerf□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□10□□□□□□□□□Java□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□Java Programming for Spatial Sciences□(CC)□□□□□□□□□□□□□□□□□□□□□□□

Matt Wood□Wellcome Trust Sanger Institute□□□□□□□□□

□□□□□□□□□□□□□□□□□□

Nathan Yau□□□□□□□□□□□□□□□□□□□□□□□□□□□□

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

□□New York Times□□□□□□□□□□□Yau□□□□□□□□□□□□□

□□□FlowingData(http□//flowingdata.com/□□□□□□□□□□□□

□□□□□□□□□□□□□□□□□□□



□□ 1□□1-1□ □□□□□□□□□□□□□□□□□□□□□□□□□□□

□□□□□□□□□□□□□□□□□□□□□□□□□□□□□GPS□□□□□□□

□□□□□□□□□(aea circle)□□□□□□□□□□□□□□GPS□□□□□

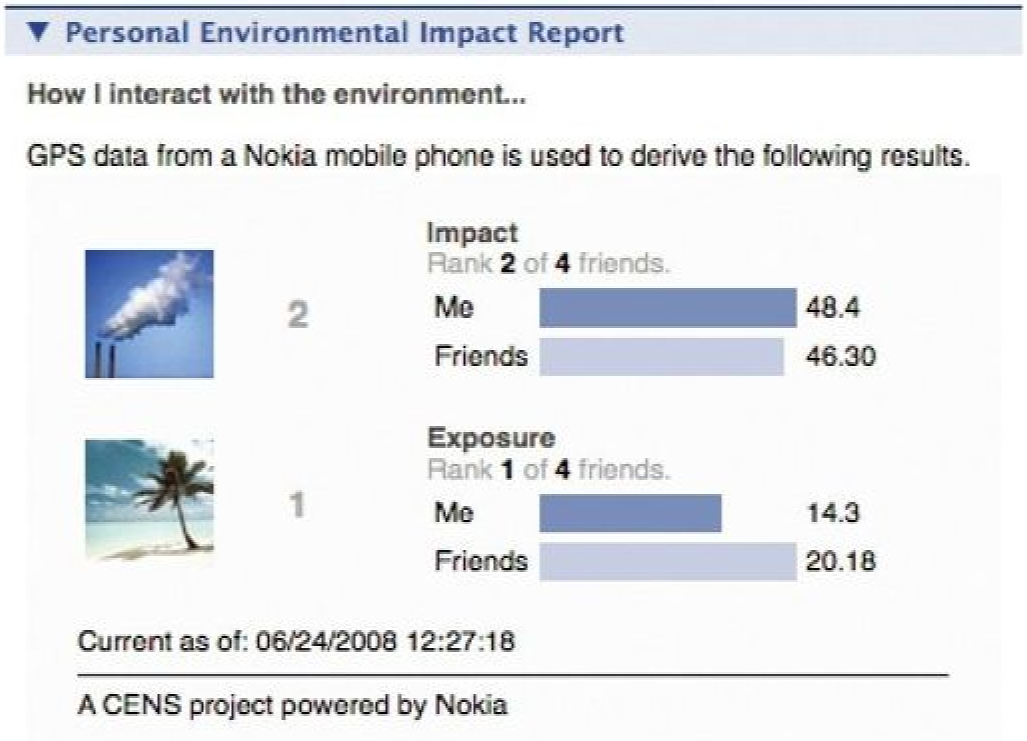
□□□□□□□□□□□□□□□□



□□　2□□1-2□　□□□□□□□□□□□□□□□□□□□□□□□□□□□□

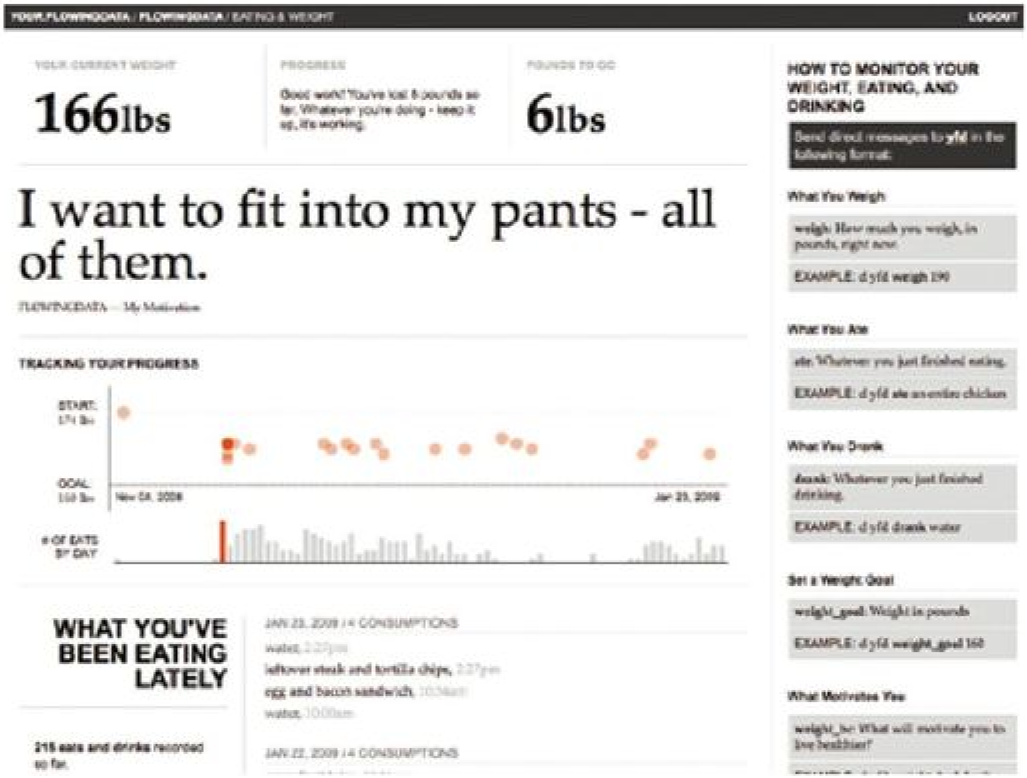
□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

GPS□□□□□□□□□



□□　3□□1-4□　PEIR□Facebook□□□□□□□□□□□□□□□□□□

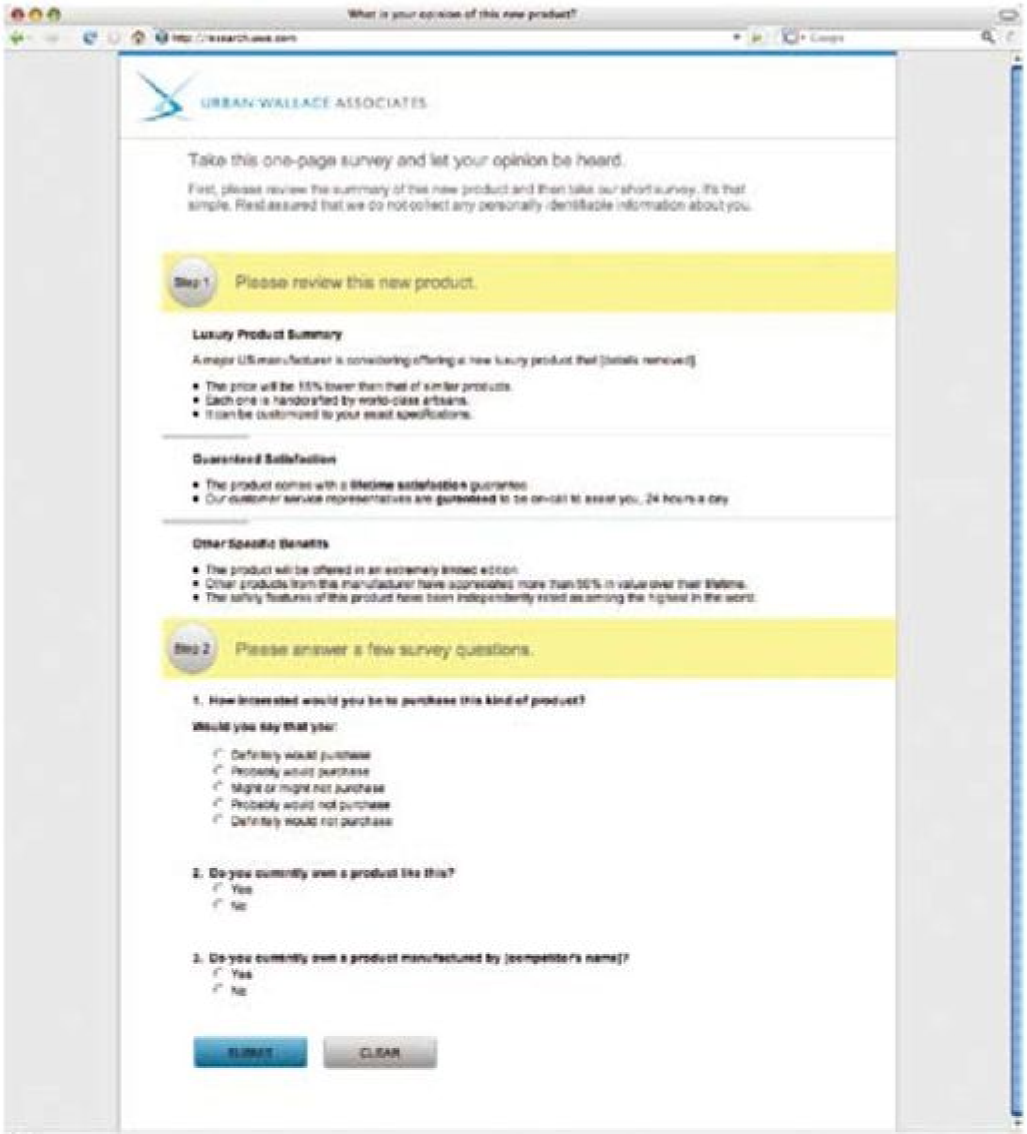
□□□□□□


LOGOUT
166lbs
6lbs
HOW TO MONITOR YOUR WEIGHT, EATING, AND DRINKING
I want to fit into my pants - all of them.
TRACKING YOUR PROGRESS
START:
GOAL:
WHAT YOU'VE BEEN EATING LATELY
JAN 23, 2009 / 4 CONSUMPTIONS
leftover steak and tortilla chips,
egg and bacon sandwich,
JAN 22, 2009 / 4 CONSUMPTIONS

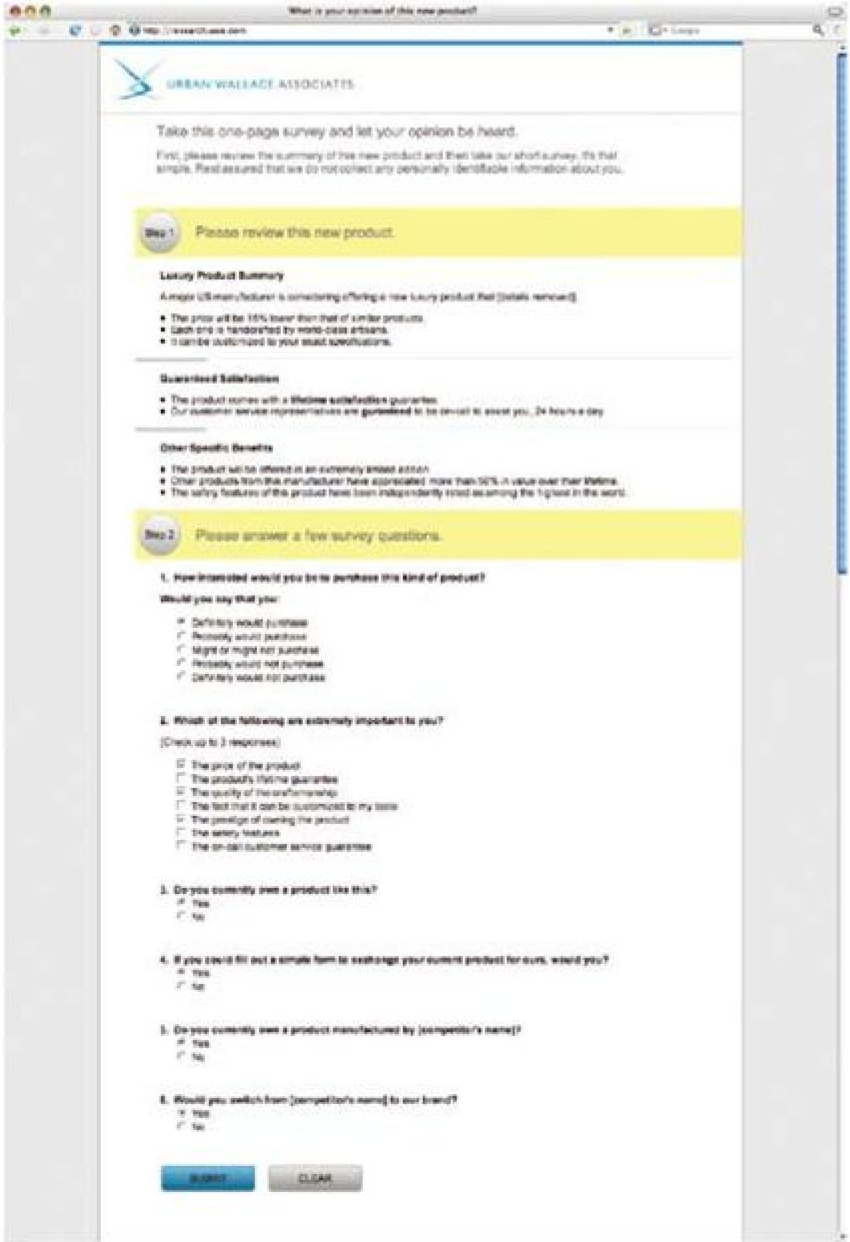
□□　4□□1-5□　□□□□□□□□□□□□□□□□□□□YFD□□□□□□□□□□□□□□□


URBAN WALLACE ASSOCIATES
Take this one-page survey and let your opinion be heard.
Please review this new product.
为了易读性，标题以Arial 20号字体显示
正文以Arial 14号字体显示
编号也把文本分成易读的分块
空白总是受人欢迎，它使页面看起来亮堂，不会让人感到压抑

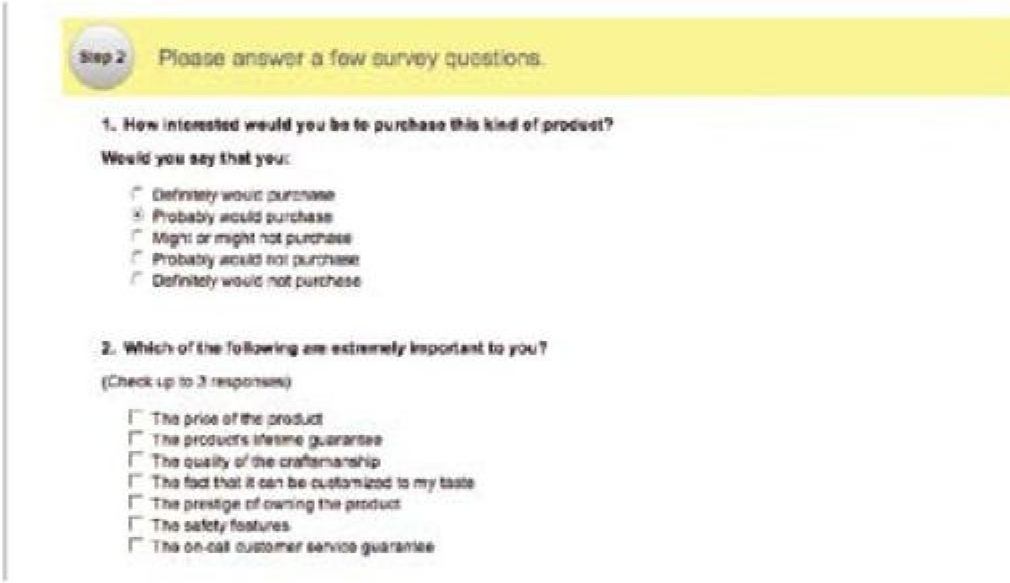
□□　5□□2-2□　□□□□□□□


URBAN WALLACE ASSOCIATES
Take this one-page survey and let your opinion be heard.
Please review this new product.
Please answer a few survey questions.
Definitely would purchase
Might or might not purchase
Yes
No
CLEAR

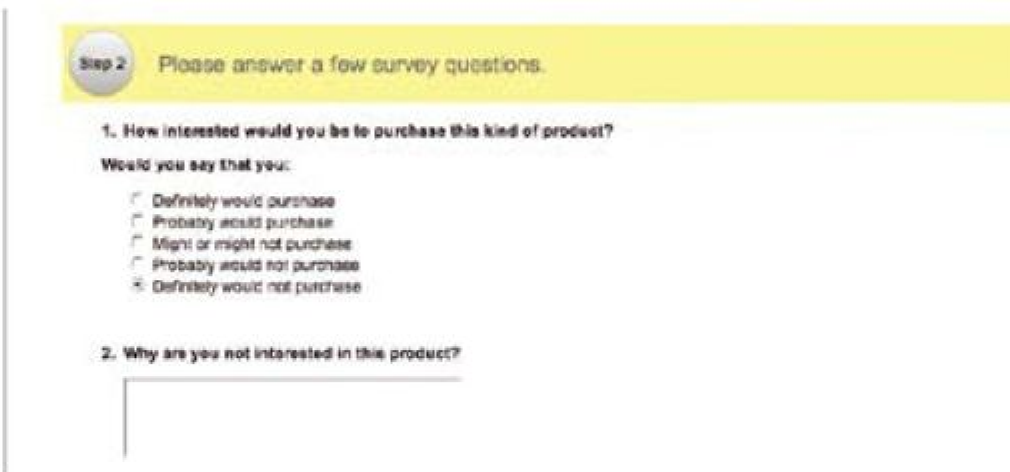
□□　6□□2-3□　□□□□□3□□□


URBAN WALLACE ASSOCIATES
Take this one-page survey and let your opinion be heard.
Please review this new product.
Please answer a few survey questions.
CLEAR

□□ 7□□2-4□ □□□□□□□□□□□□□□□□6□□□



□□ 8□□2-5□ □□□□——□□□□□1□□“Yes”□

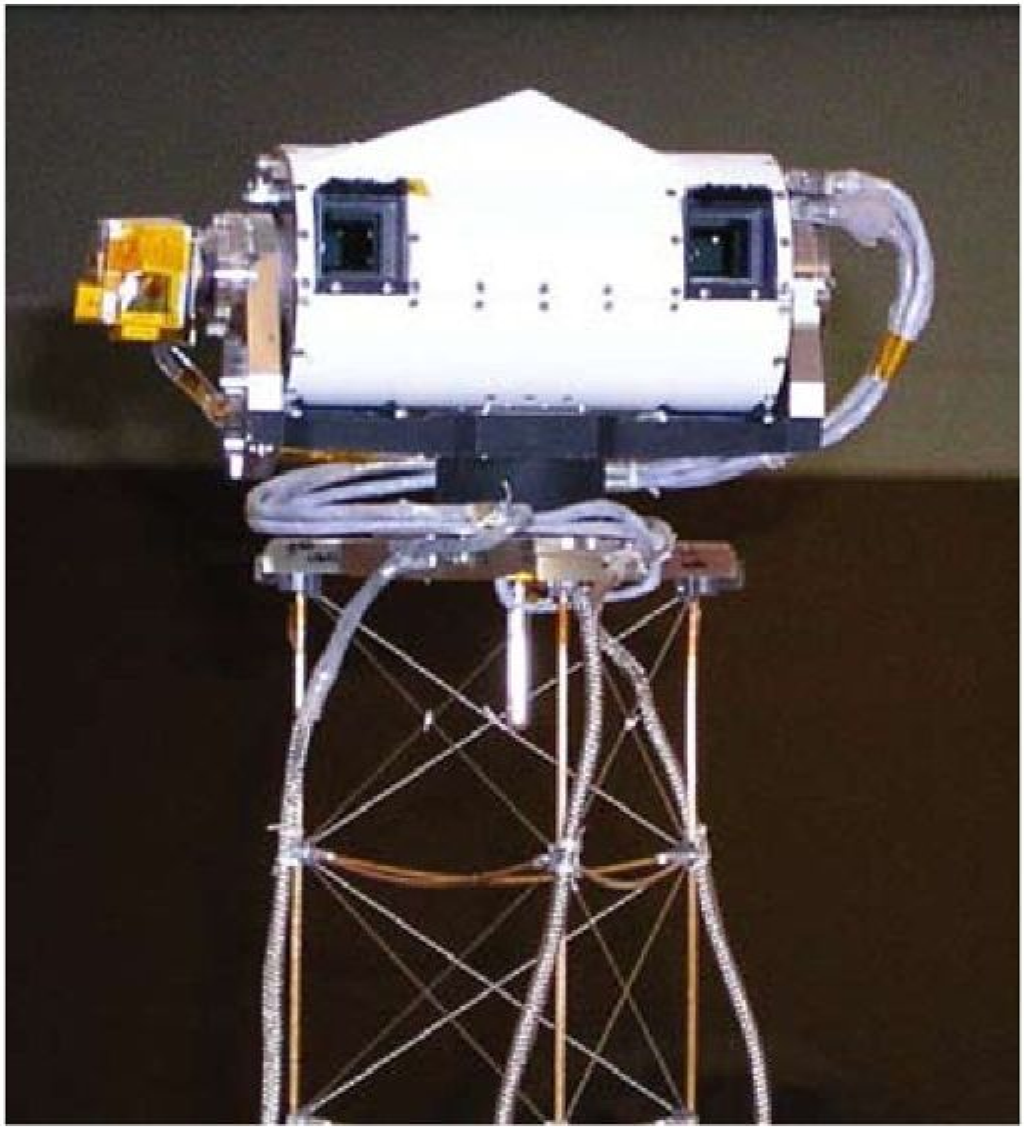


□□ 9□□2-6□ □□□□——□□□□□1□□“No”□



□□ 10□□3-1□ □□□□□□□□□□□□□□□□□□□□□□□□□□□□

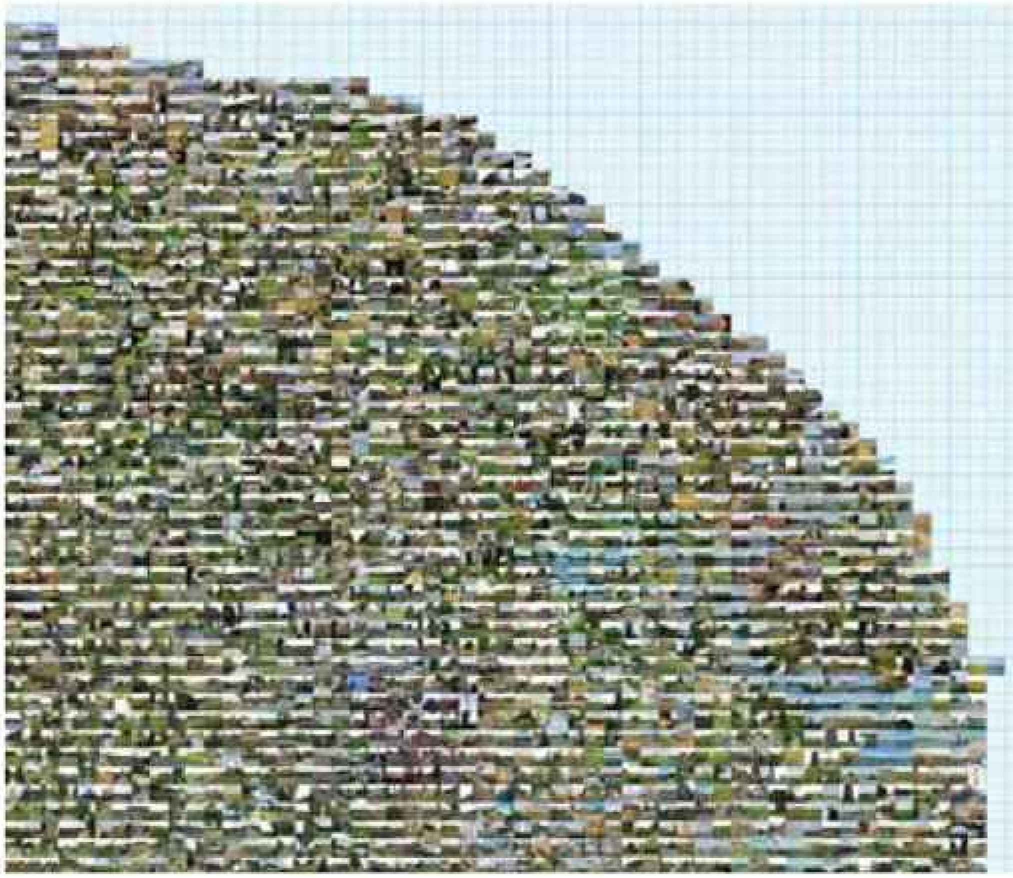
□□NASA/JPL)



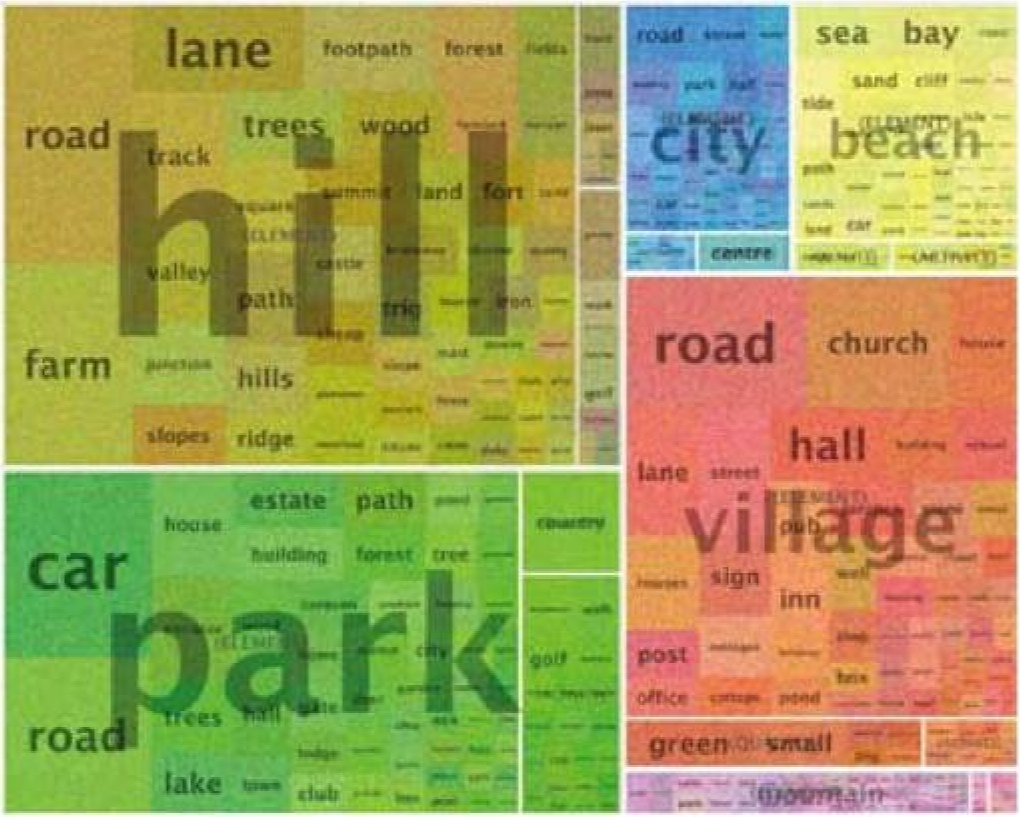
[illegible] 11[illegible]3-2[illegible] [illegible]Arizona[illegible]/NASA/JPL)



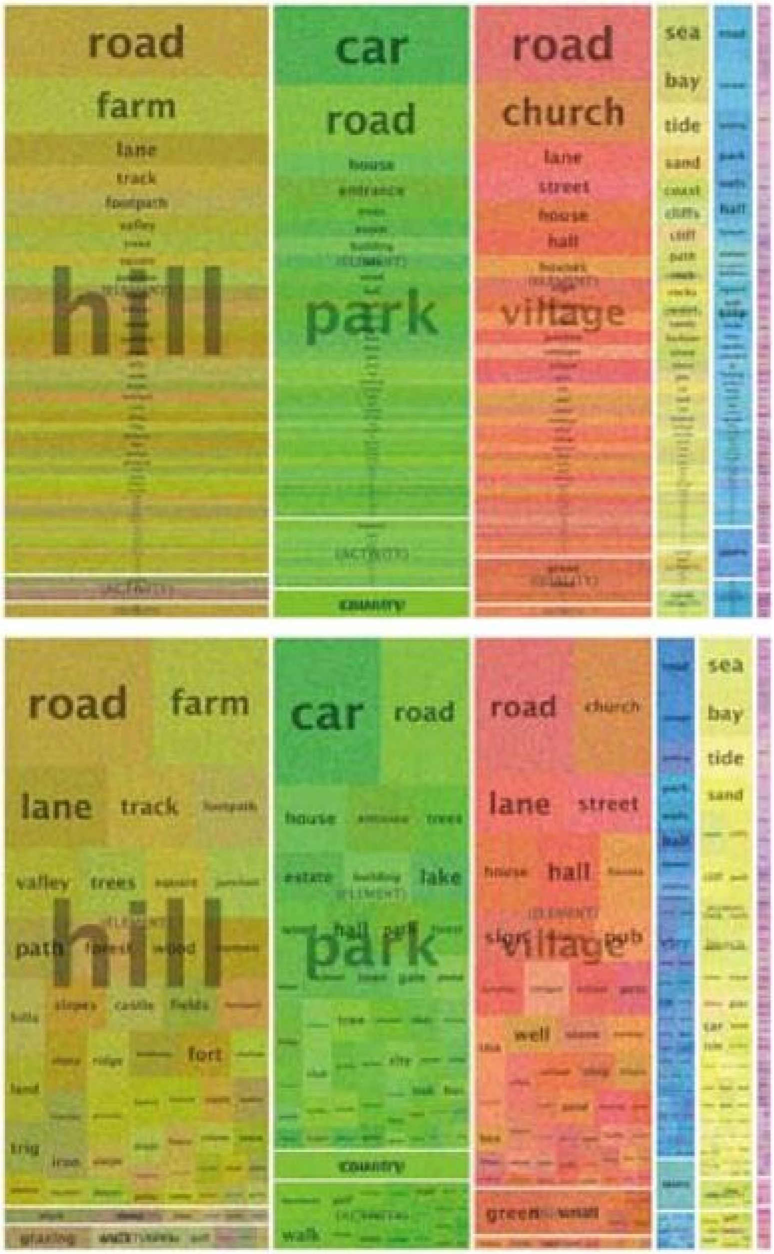
□□ 12□□6-1□ Aberuchill□□□□□□□□□Bioran Dalchonzie□□□ Geograph □□□□□□□□□ 100 □□□□(htp □//www.geograph.org.uk/photo/1006884 □□□□□□□ Dr.Richard Murray □□□□ Creative Commons License(http□//creativecommons.org/licenses/by-sa/2.0/□□□□□



□□ 13□□6-2□ Norfolk□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□ 1□□□□□□□□□□□ Creative Commons License(http□//creativecommons.org/licenses/by-sa/2.0□□□□□


hill
lane
footpath
forest
road
track
trees
wood
summit
land
fort
valley
castle
path
trig
iron
farm
junction
hills
slopes
ridge
car
park
estate
path
house
building
forest
tree
country
road
trees
hall
lake
town
club
golf
road
city
centre
sea
bay
sand
cliff
beach
tide
car
road
church
hall
lane
street
village
pub
sign
inn
post
office
cottage
pond
green
small

[illegible] 14[illegible]6-4[illegible] [illegible]/[illegible](fcet)/[illegible]“[illegible]”[illegible]


road
farm
lane
track
footpath
valley
hill
car
road
house
entrance
park
country
road
church
lane
street
house
hall
village
sea
bay
tide
sand
coast
cliffs
road
farm
lane
track
valley
trees
path
forest
wood
slopes
castle
fields
fort
hills
land
trig
iron
grazing
car
road
house
trees
estate
lake
hall
park
country
walk
road
church
lane
street
hall
village
well
green
sea
bay
tide
sand

□□ 15□□6-5□ □□□□□□□□□□□□□□□□□□□□□□□□

□□□□□□□□□□□□□□□□□□□□□/□□/□□□□□□□“□□□□□”□

□□□□□□□□□□□“□□□□/□□□□□”□□□□□□□□□□□□


lane
road
trees
wood
hill
farm
hills
ridge
path
car
park
road
estate
path
trees
lake
sea
bay
beach
city
road
church
hall
village
lane
sign
inn
post
office
green
small

[illegible] 16[illegible]6-6[illegible] “[illegible]”[illegible]CIELab[illegible]

[illegible]


hill
park
village
road
lane
car
farm
road
beach
city
road
church
hall
sea
bay
trees
wood
estate
path
green
small
lake
inn
hills
ridge
path

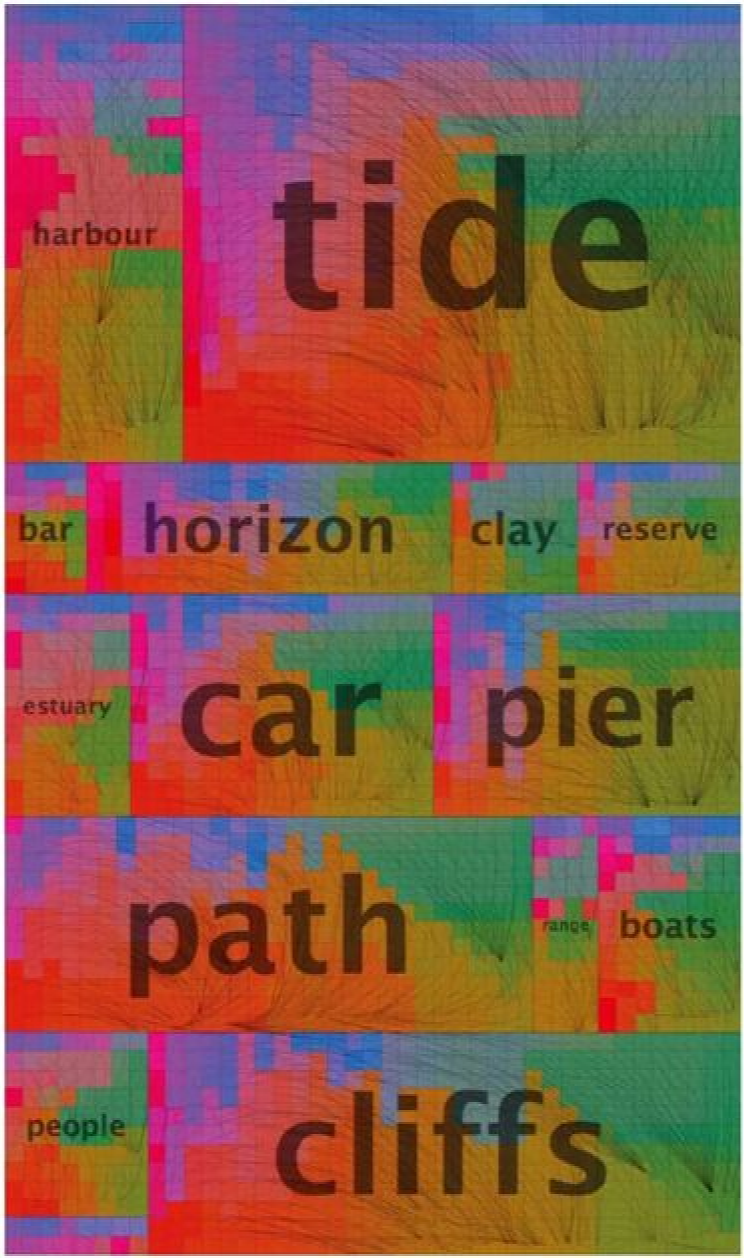
[illegible] 18[illegible]6-8[illegible] [illegible]CIELab[illegible]


farm
city
track
hill
forest
trees
wood
road
lane
green
houses
hall
street
house
village
road
lane
bay
sea
tide
hall
forest
lake
wood
park
car
road
trees
church
sign

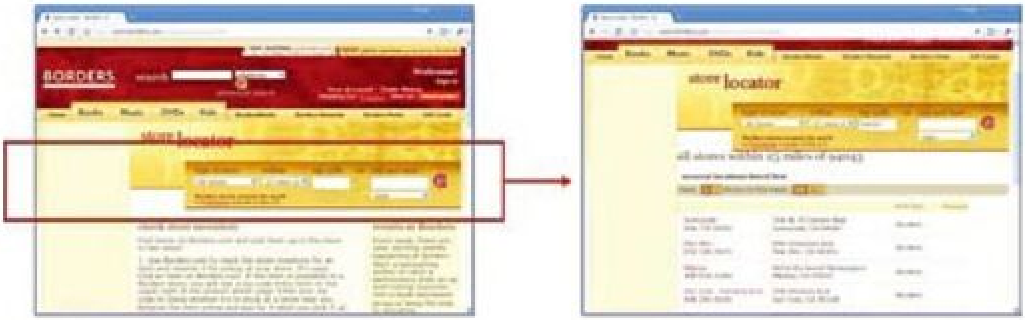
□□ 17□□6-7□ “□□□□□□”□□□□□CIELab□□□□□□□□□□□□□
□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□


farm
city
track
hill
wood
green
houses
hall
road
street
lane
house
village
road
lane
hall
forest
lake
park
trees
church

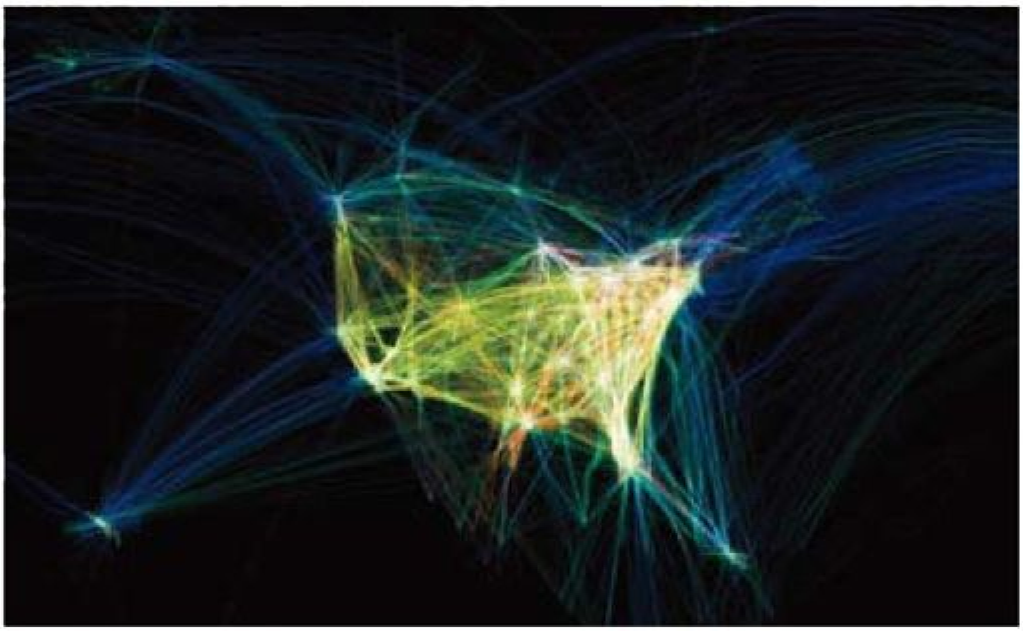
[illegible] 19 [illegible] 6-9 [illegible]



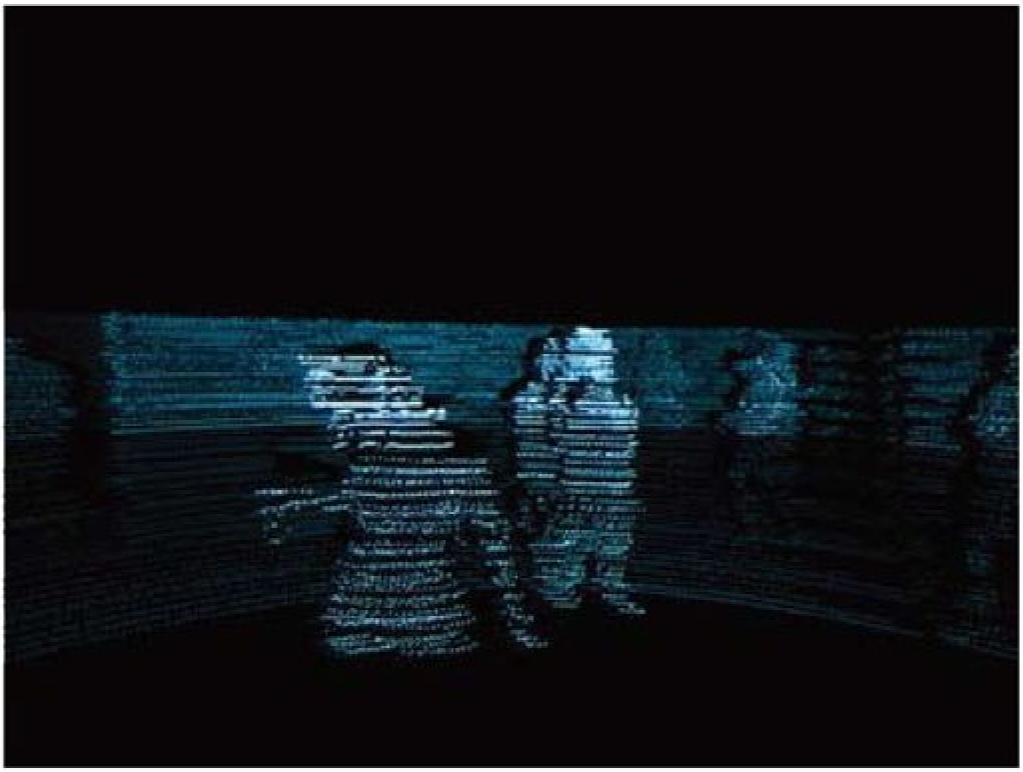
□□ 20□□6-10□ □beach□□□6-9□□□□□□□□□□□□□□□□□



□□ 21□□9-1□ Borders□□□□□□□□□□□□□□□□Deep Web□□□

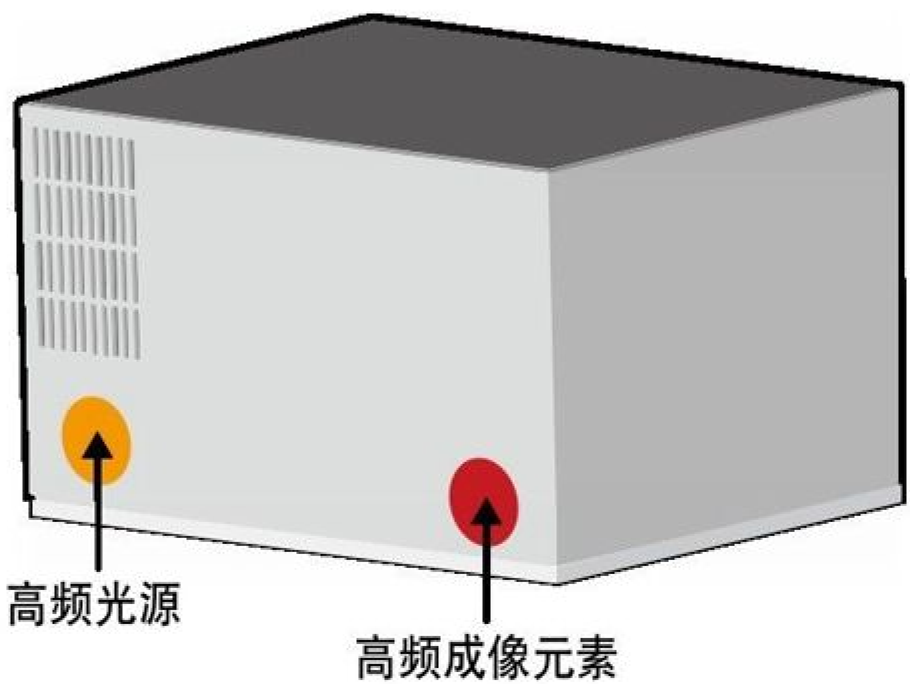


□□ 22□□10-1□ □□“□□□□”□□□□□□2005□

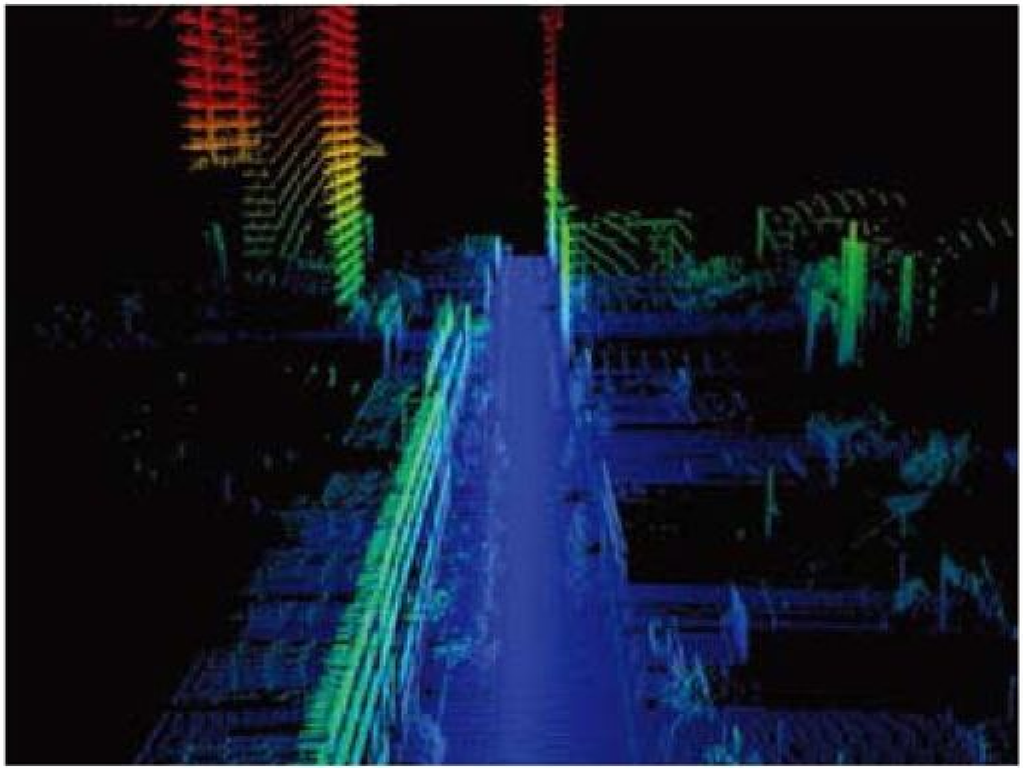


[illegible] 23[illegible]10-3[illegible] [illegible]Velodyne[illegible]Lidar[illegible]

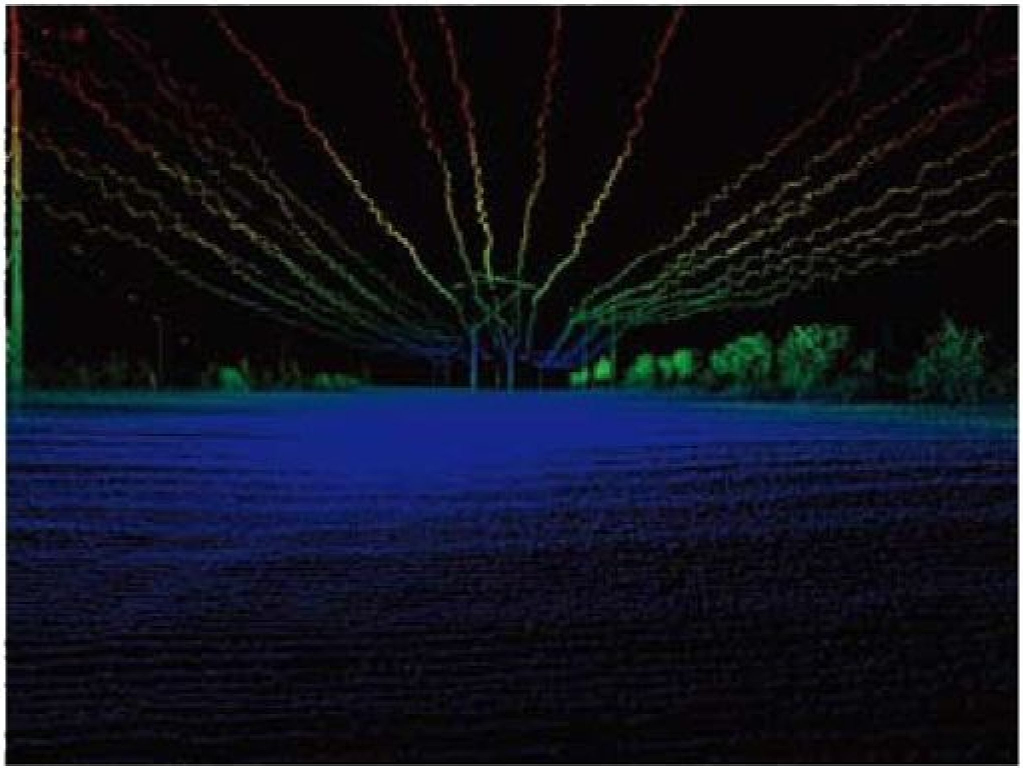
[illegible]



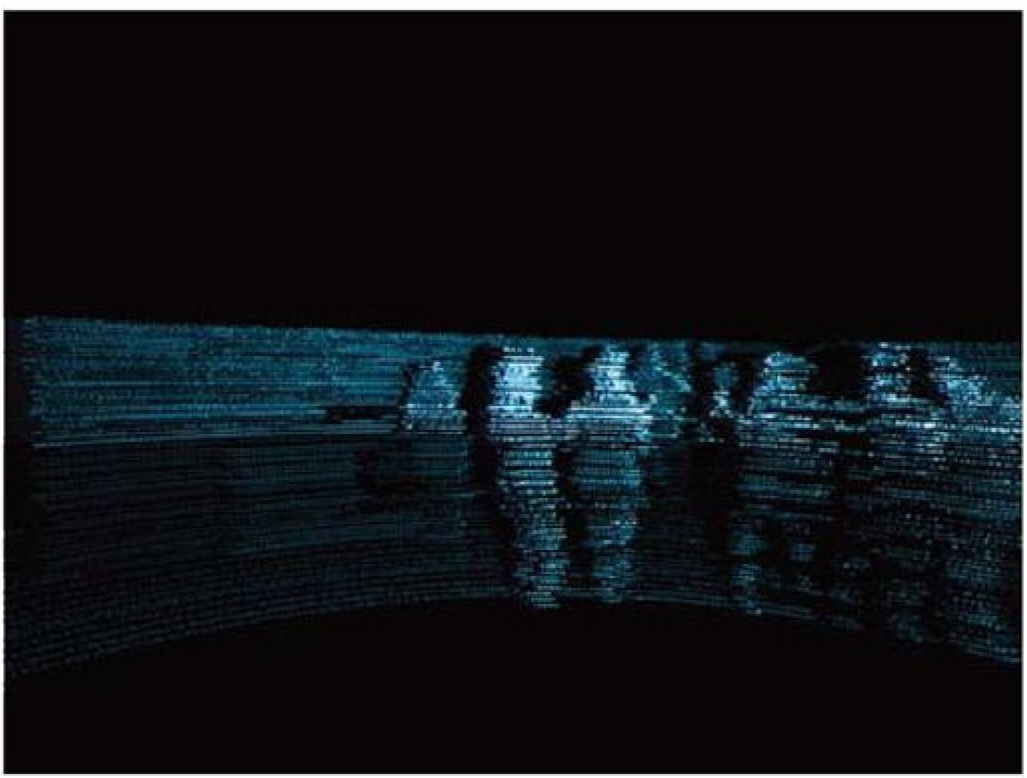
□□ 24□□10-4□ Geometric Informatics□□□□□□□□Geometric Informatics□□□



□□ 25□□10-5□ Velodyne Lidar□□□□□□□□□□□□



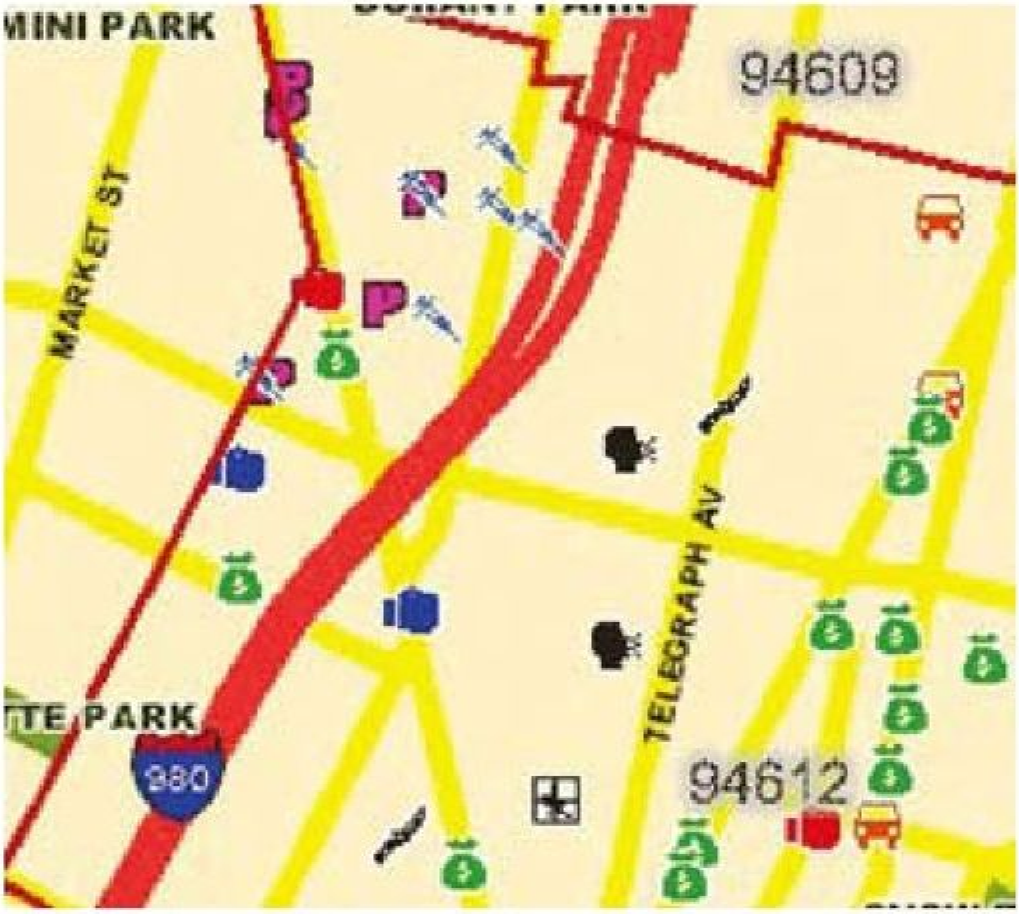
□□ 26□□10-6□ Velodyne Lidar□□□□□□□□□□□



□□ 27□□10-7□ □□□□□□□□□□□□Velodyne Lidar□□□□□□64□□□□□□□Lidar□□□□□□□□



□□ 28□□10-8□ □□Thom Yorke□Radiohead□□□□□□□□□□



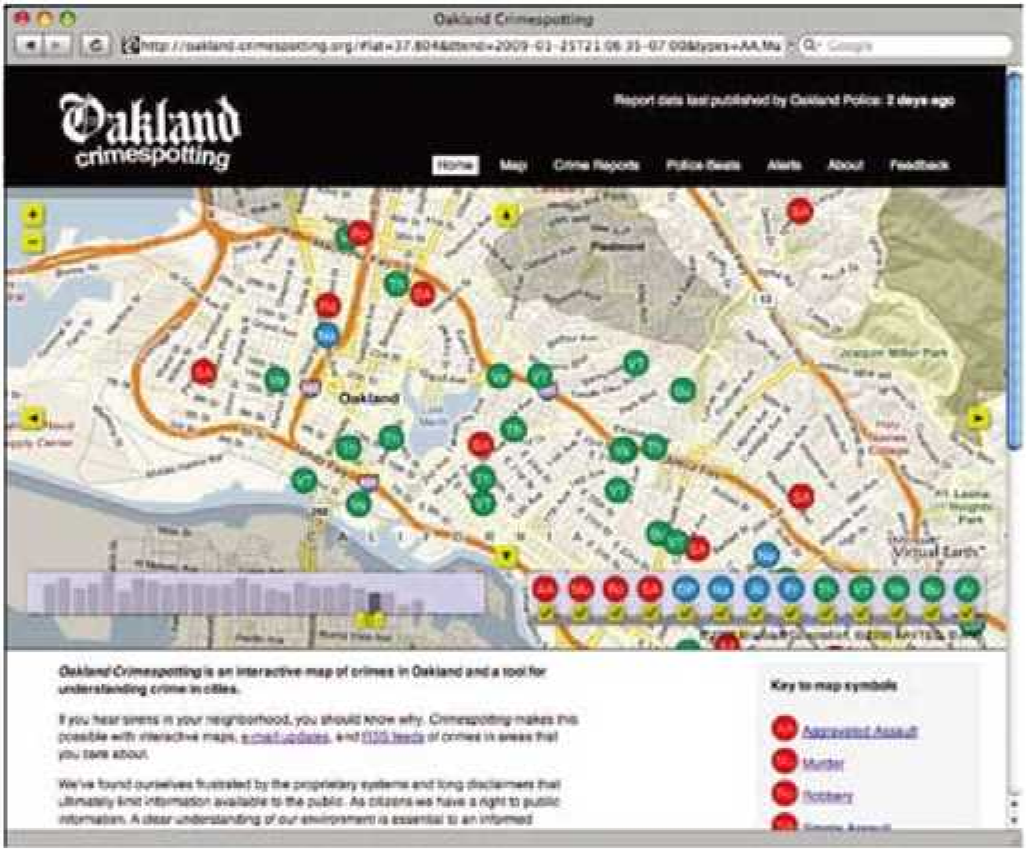
29 11-2 CrimeWatch



□□ 30□□11-3□ CrimeWatch□□□□□□□□□□□□□□□□□□□□□



□□ 31□□11-5□ □□□□□□□□□□□□□□□□□□□□□□□



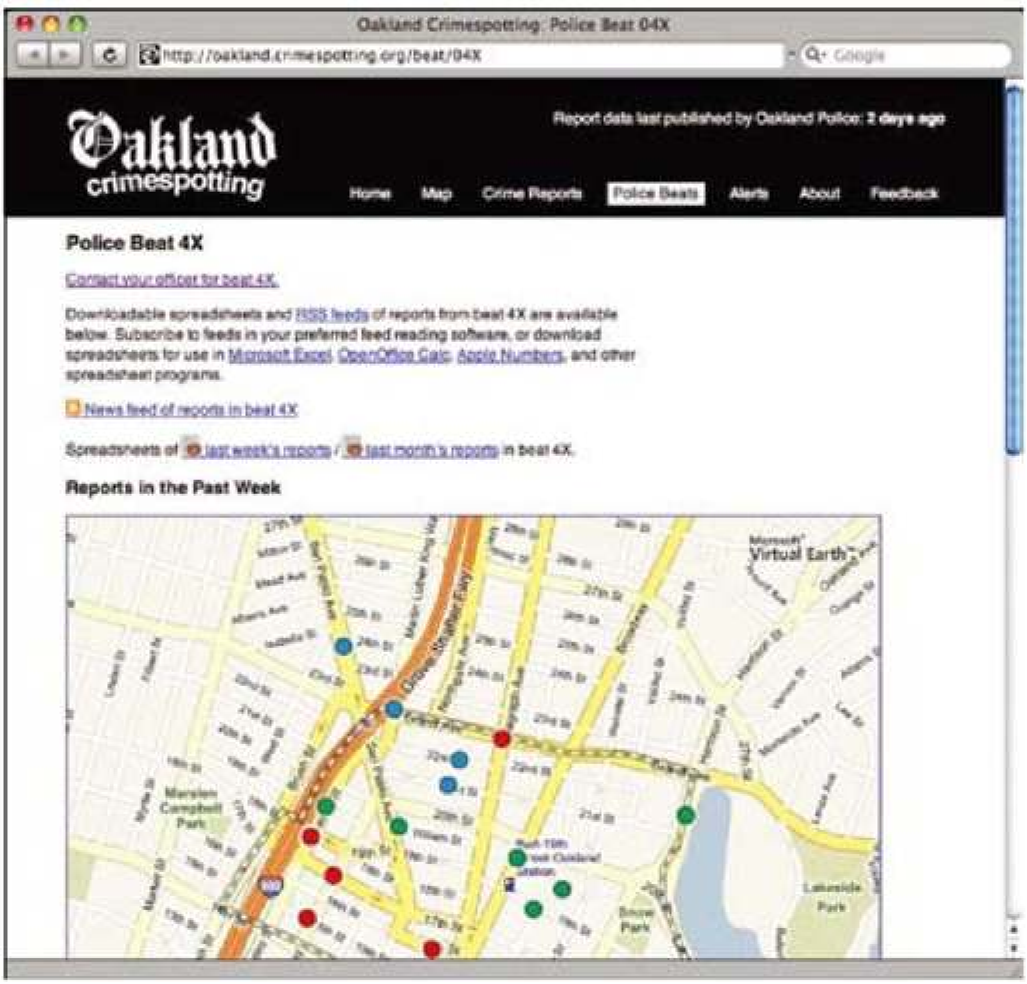
□□ 32□□11-6□ □□□Crimespotting□□□□□□□□□□□□□□□□□□□□□



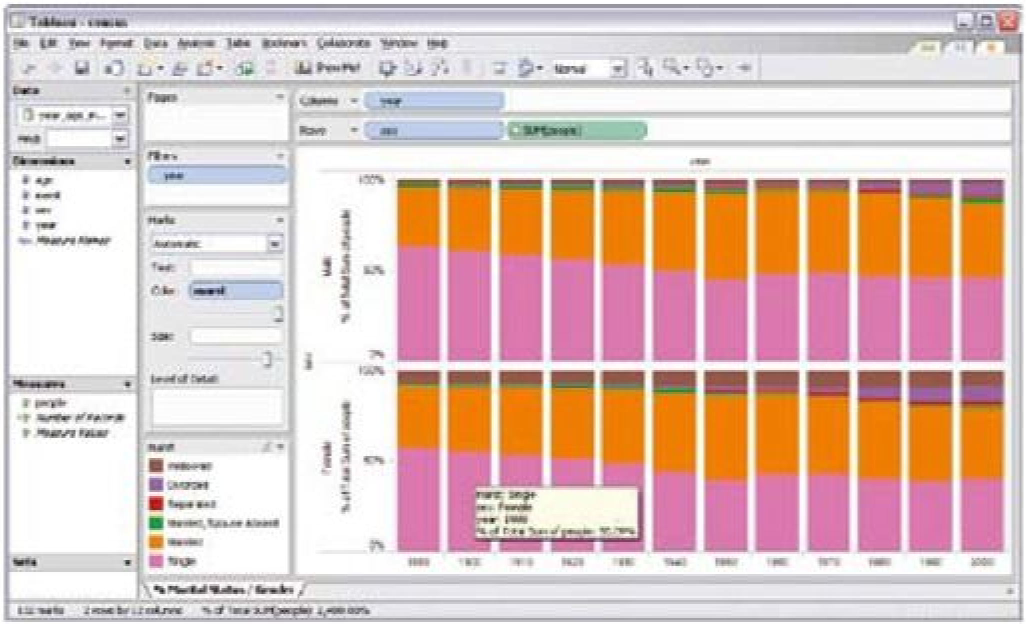
□□ 33□□11-7□ Crimespotting□□□□□□□□□□□□□



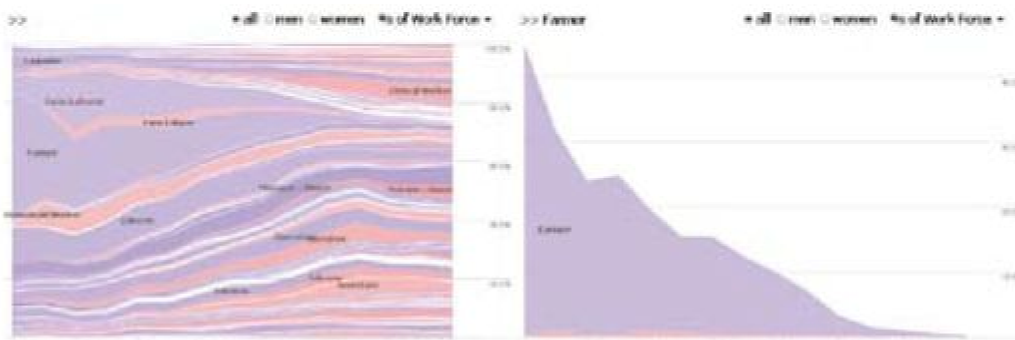
□□ 34□□11-8□ □□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□



□□ 35□□11-10□ □□□□□□□□□□□□□□□□□□□□□□□□□□□□□

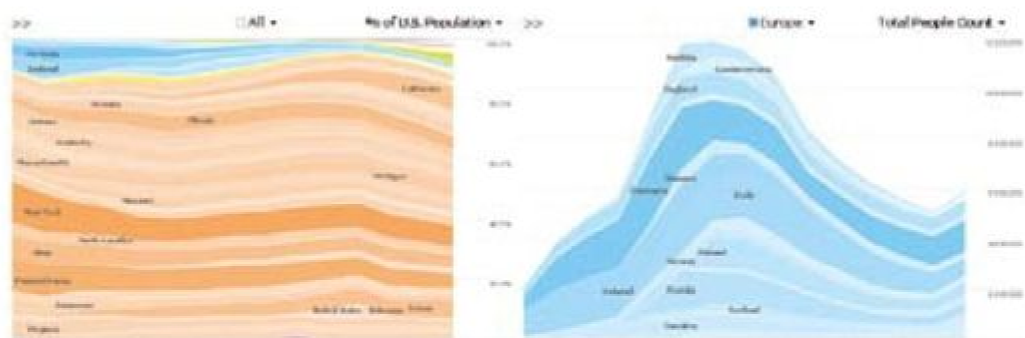


□□ 36□□12-3□ □□Tableau□□□□□□□□□□□□□□□□□□□□



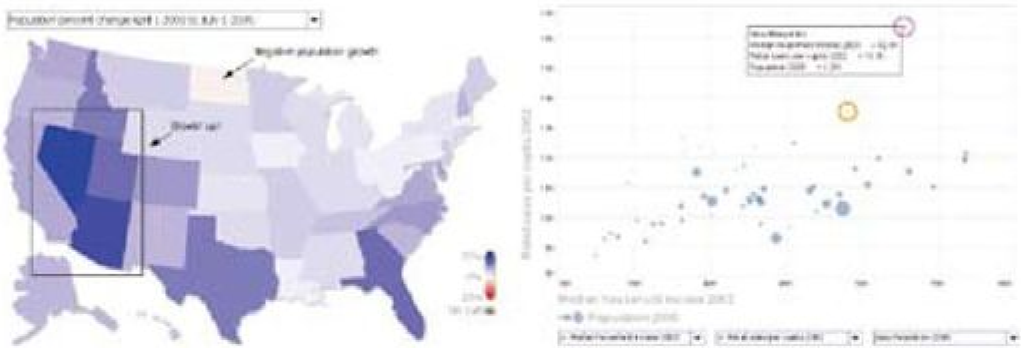
□□ 37□□12-4□ □□□□□□□□□□□□□150□□□□□□□□□□□□

□□□□□□□□□□□

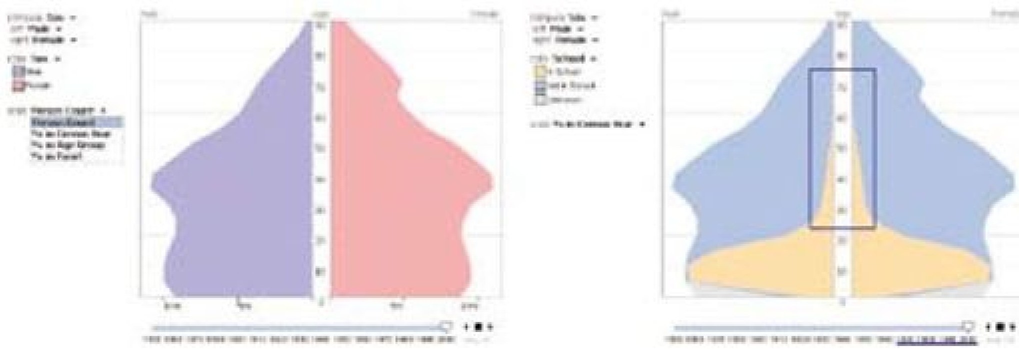


□□ 38□□12-5□ □□□□□□□□□□□□□150□□□□□□□□□□□

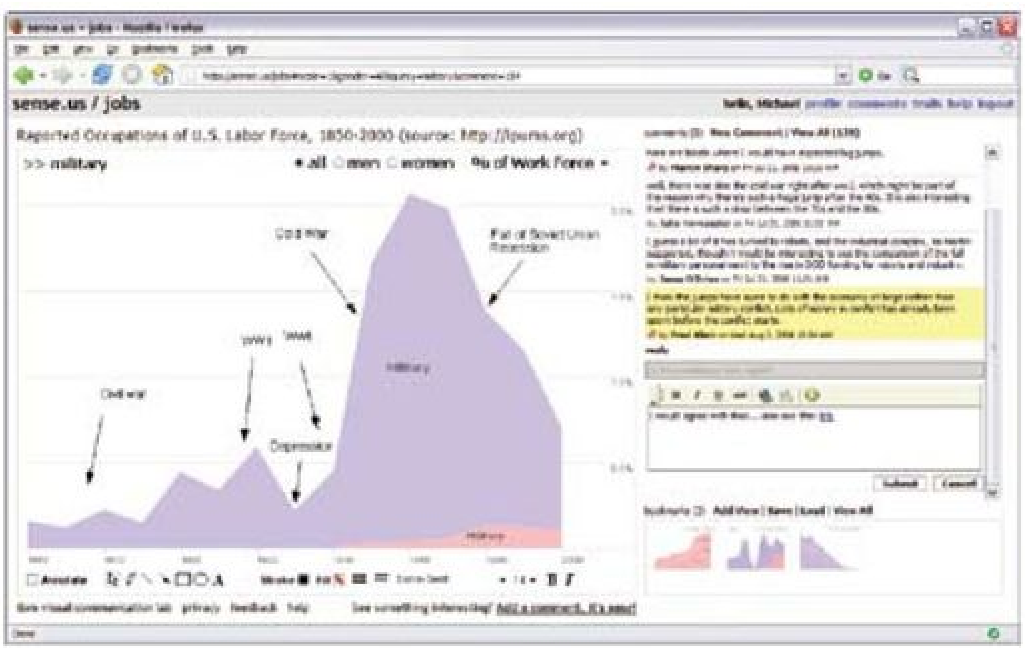
□□□□□□□□□□□□



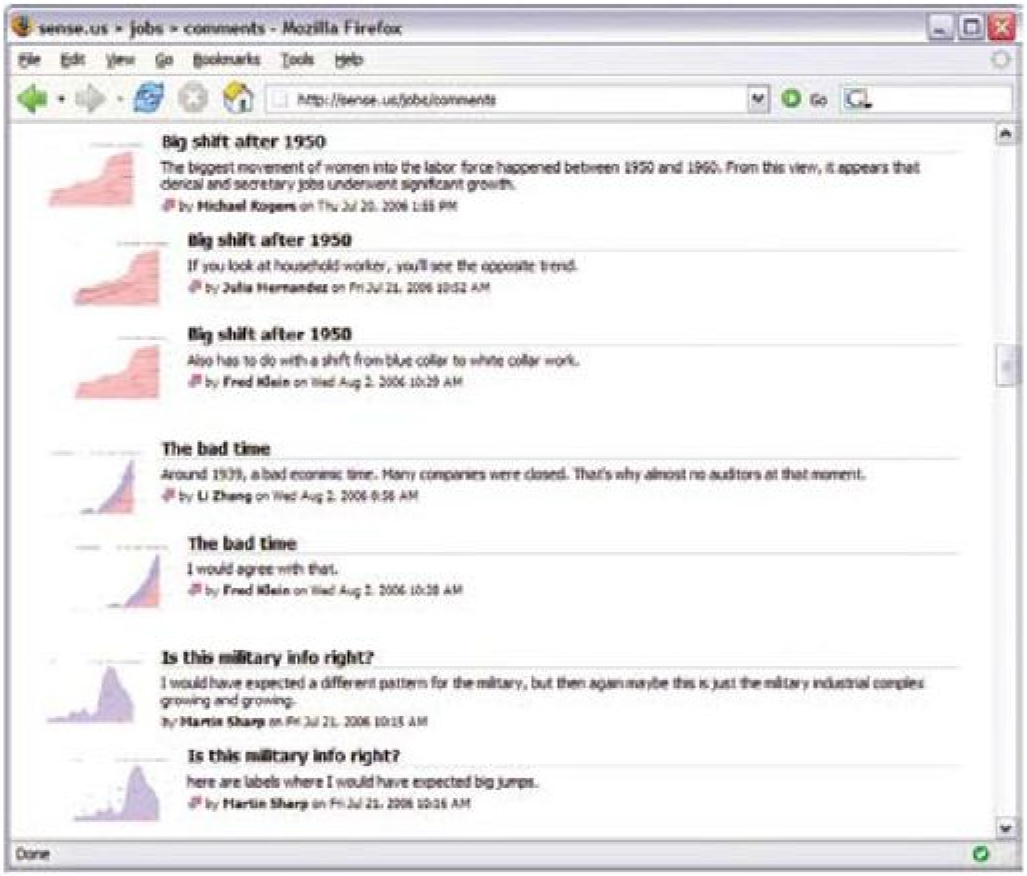
[illegible] 39[illegible]12-6[illegible] [illegible]2000[illegible]2005[illegible]
[illegible]x[illegible]y[illegible]



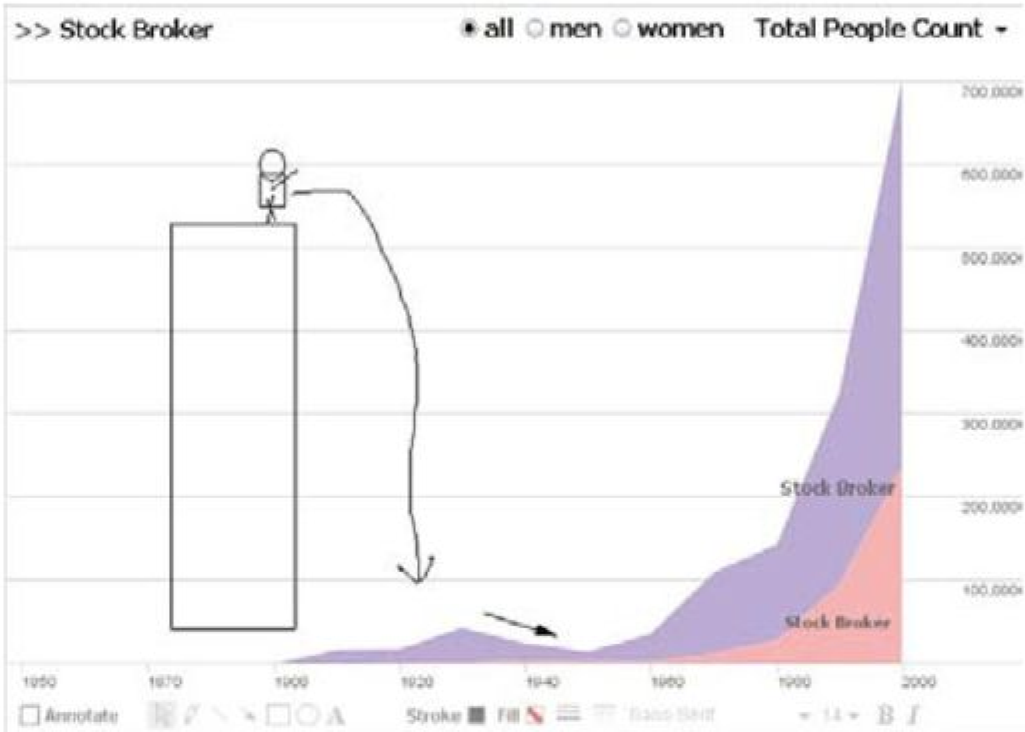
[illegible] 40[illegible]12-7[illegible] [illegible]2000[illegible]
[illegible]2000[illegible]



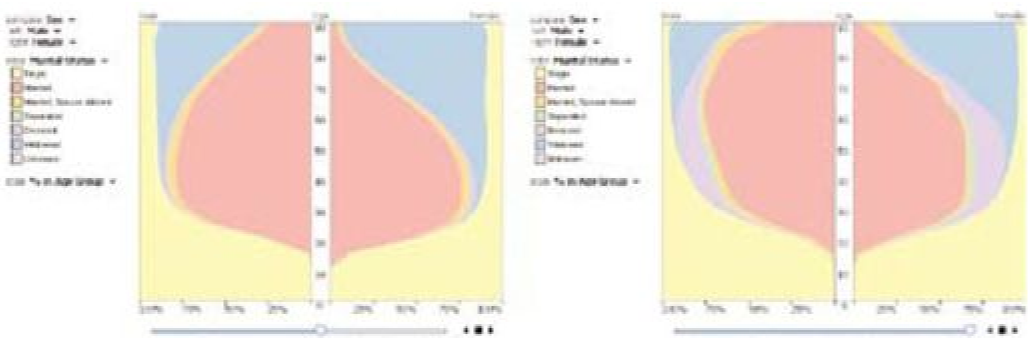
□□ 41□□12-8□ sense.us□□□□□□□□□a□□□b□□□□□□□□□□c□□□□□□□□□□d□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□e□□□□□□□□□□□□□□□□f□□□□□□□□□□URL□□URL□□□□□□□□□□□□



□□ 42□□12-9□ sense.us□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

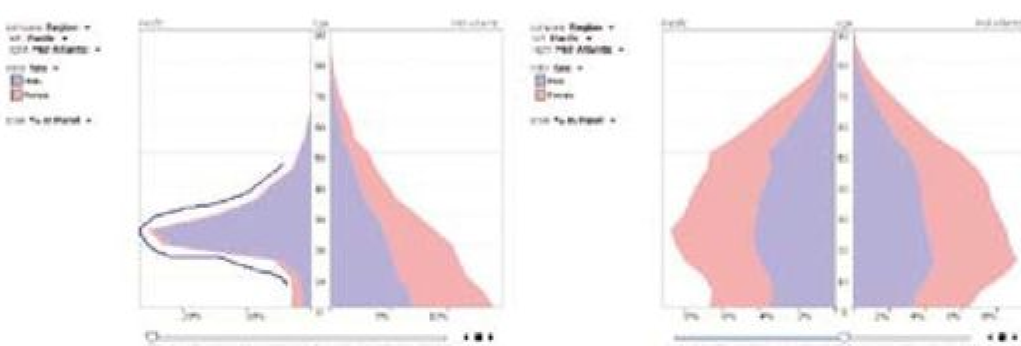


[illegible] 43[illegible]12-10[illegible] [illegible]“[illegible]‘[illegible]’[illegible]”



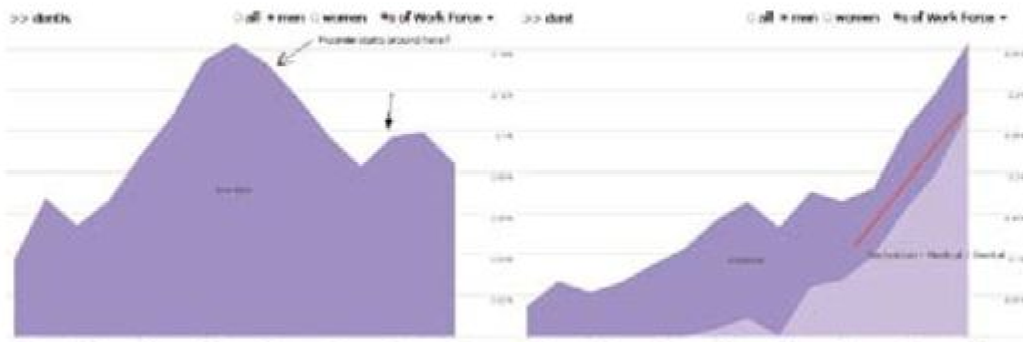
[illegible] 44[illegible]12-11[illegible] [illegible]1940[illegible]2000[illegible]

[illegible]

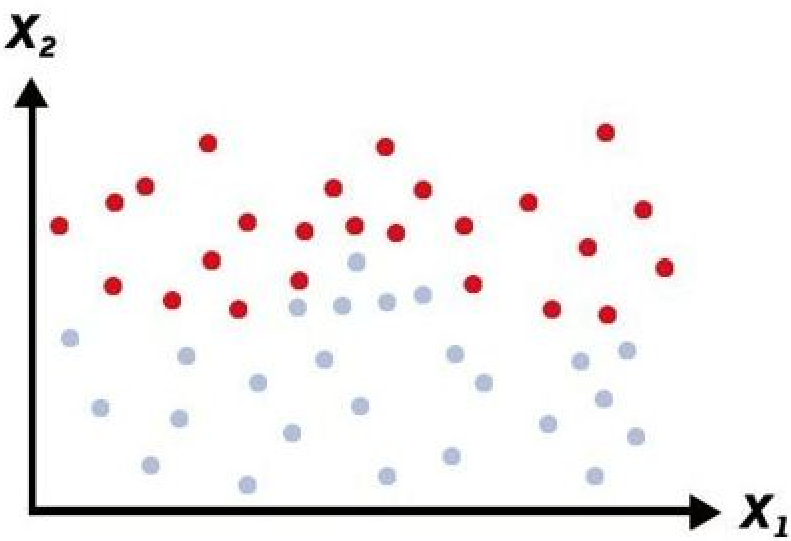


□□ 45□□12-12□ □□□□□□□□□1850□□□□□□1940□□□□□□□□□□□□

□□□□□□□



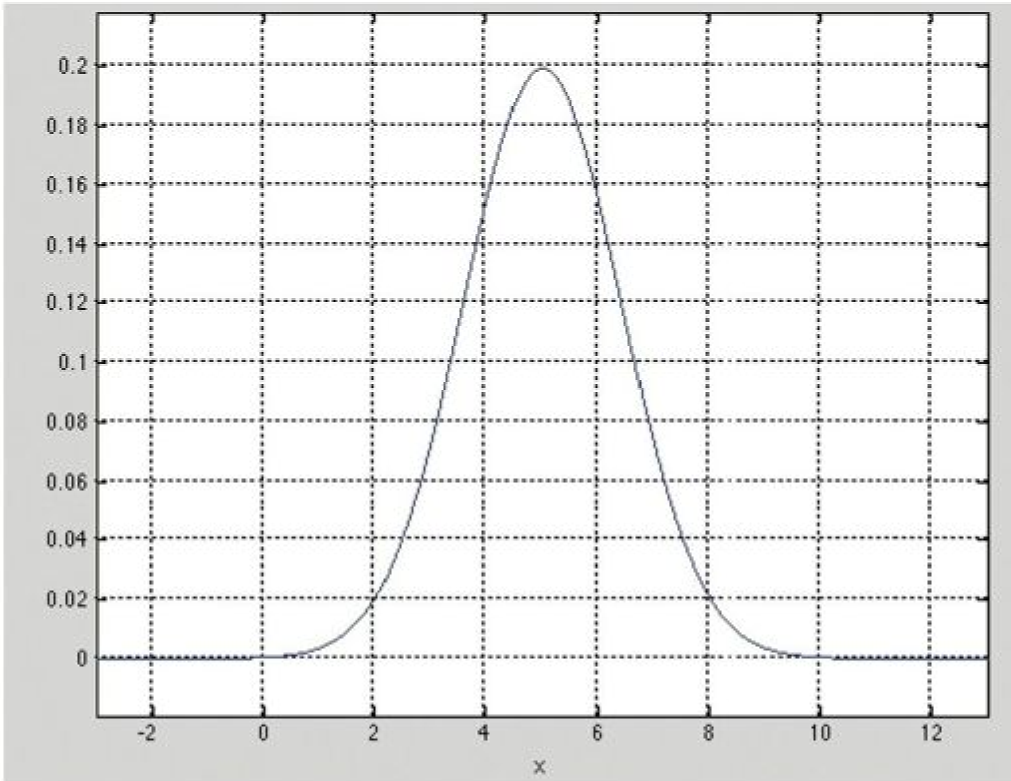
□□ 46□□12-13□ □□□□□□□□□□□□□□□□□□□1930□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□



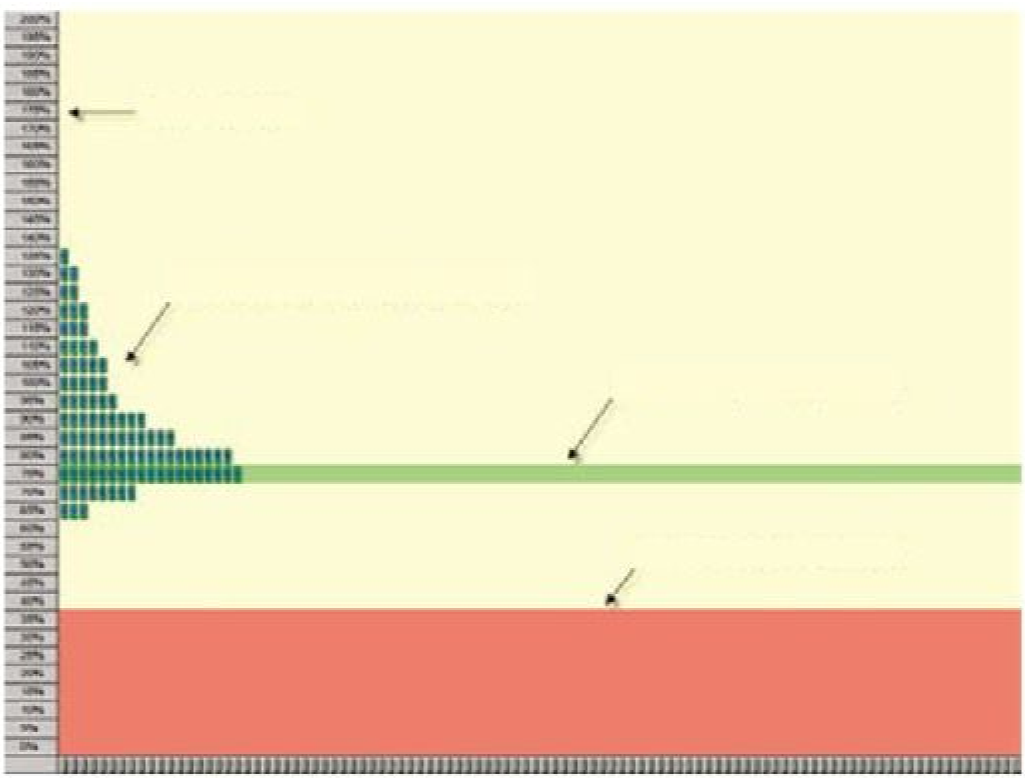
□□ 47□□13-2□ □□□□□□□□□□□□□□□□



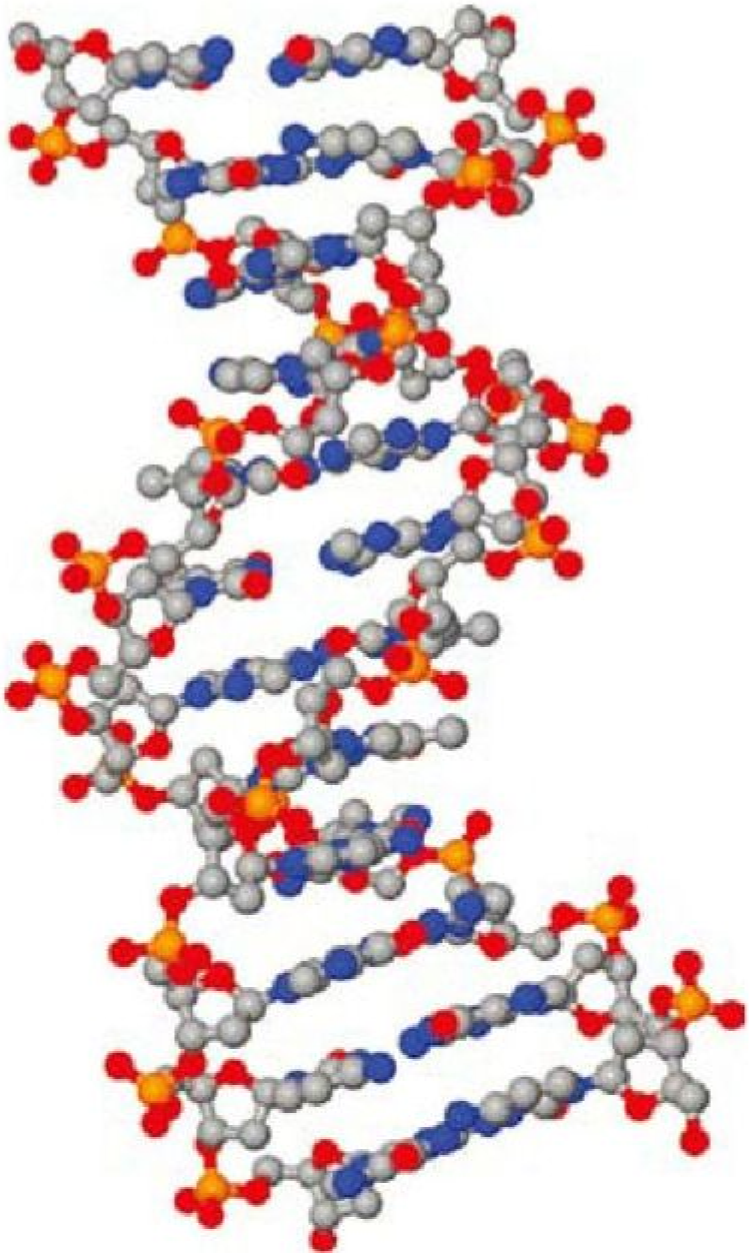
□□ 48□□13-3□ □□□□□a□b□c□□2005□□□□



□□ 49□□13-4□ □□□□

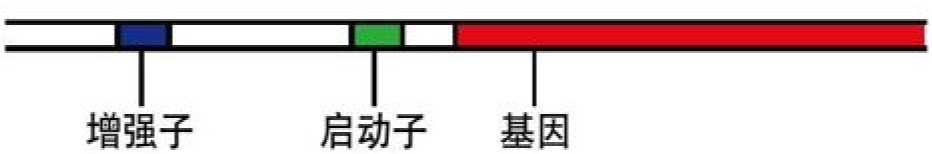


□□　50□□13-6□　Goldstein□□□□□□□□□□□□□□□□□□□□□□



□□　51□□15-1□　DNA□□□□□□□□□□POV-Ray□□□PDB□□1BNA□□□

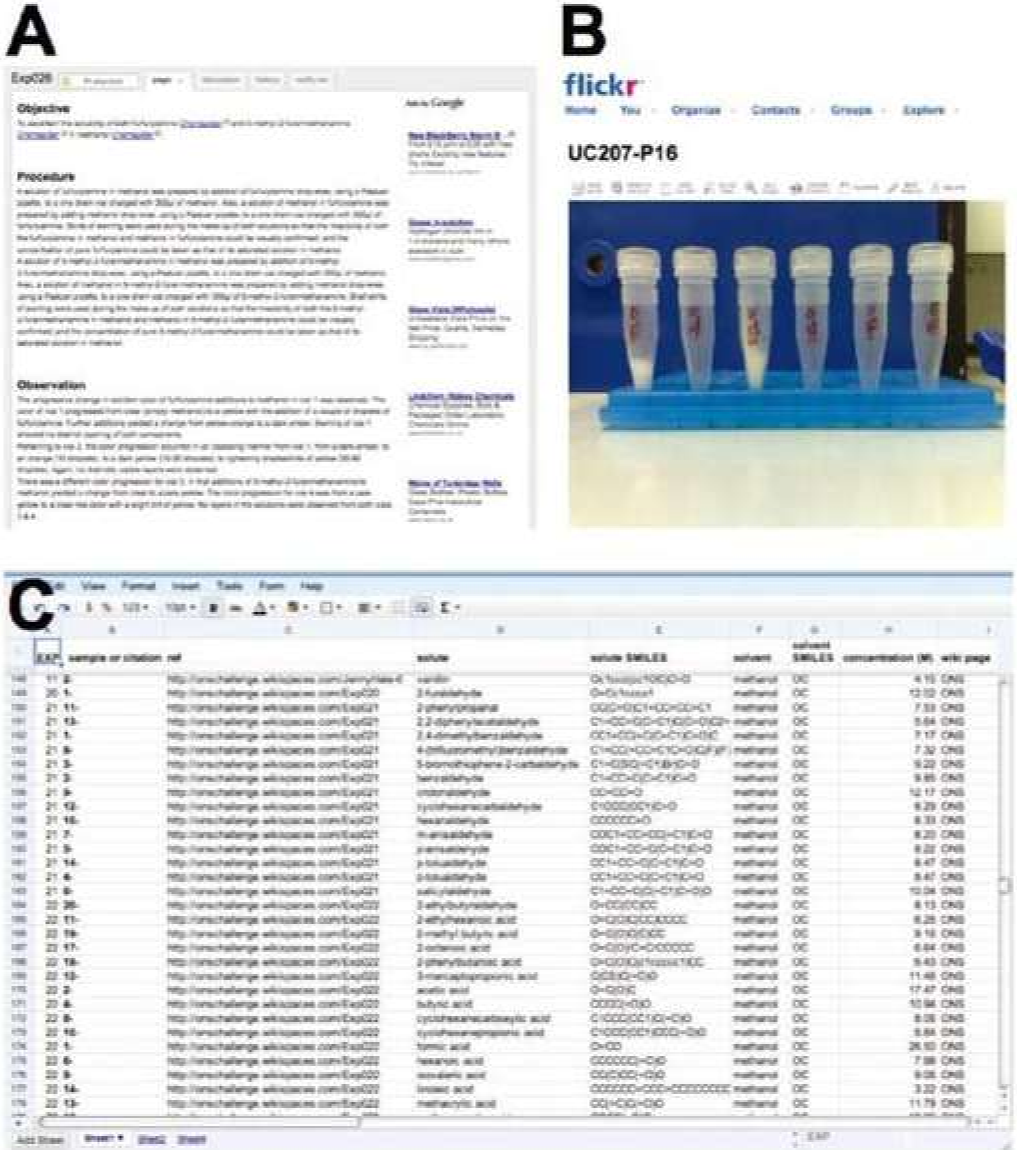
doi□10.2210/pdb1bna/pdb



□□ 52□□15-2□ □□□□DNA□□□□□□□□□□□□□□□□□□□□□

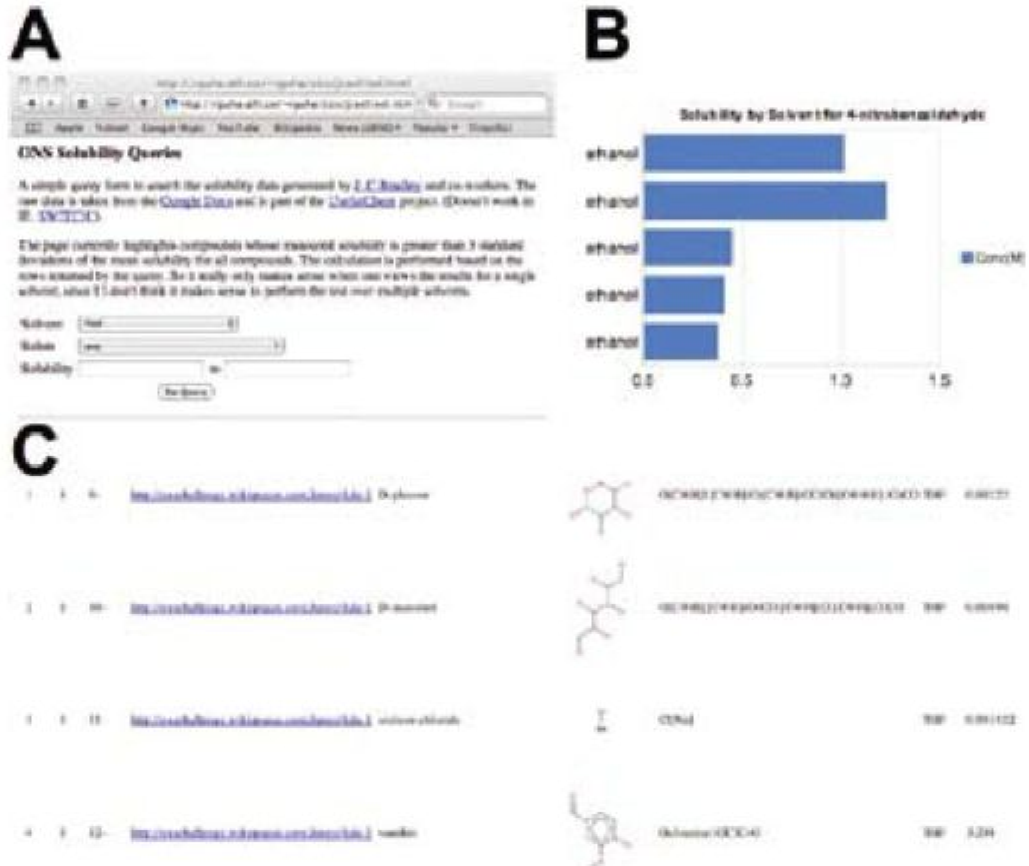
□□

□□□□□□□□□□

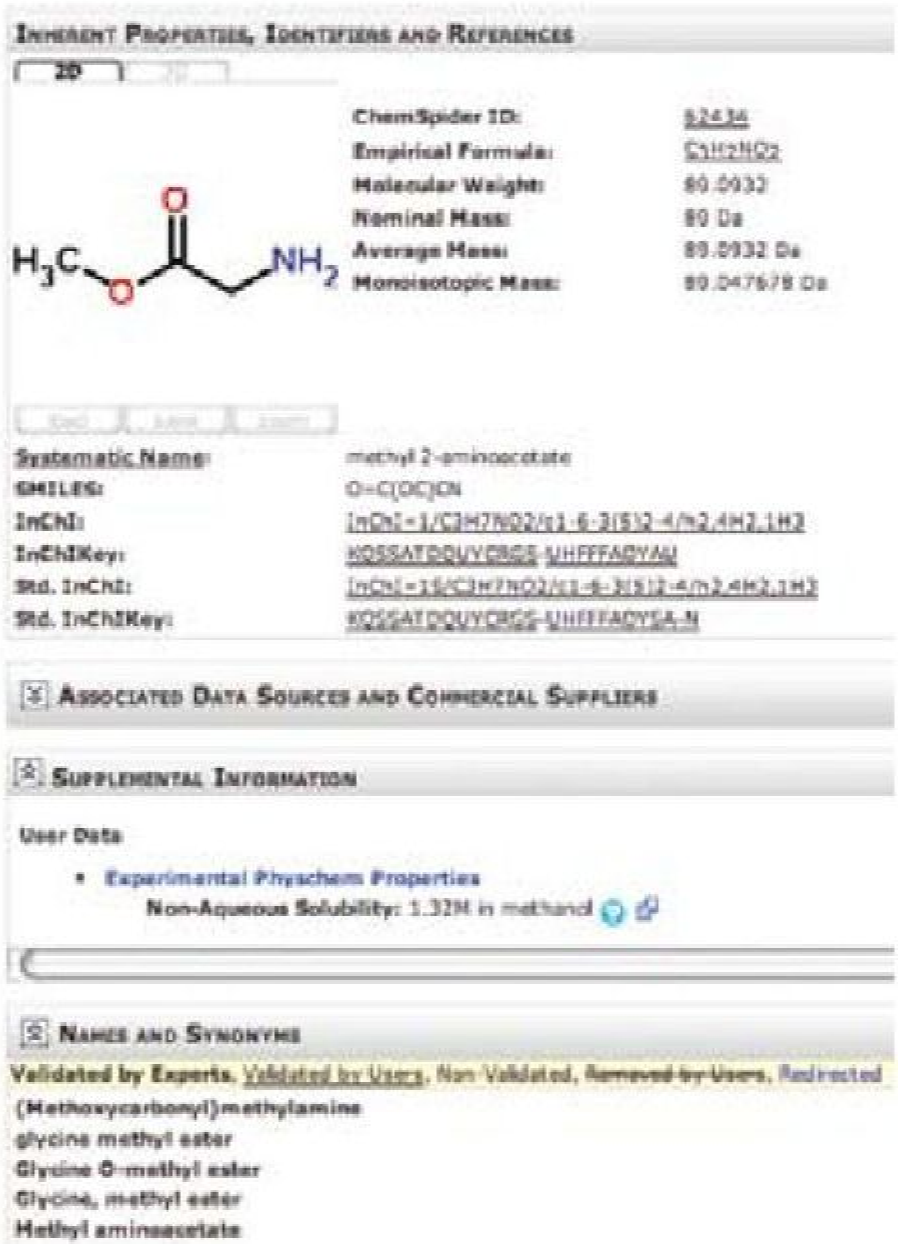


□□ 53□□16-1□ □□□□□□□□□□□□□□□□□□□□A□□□□□□

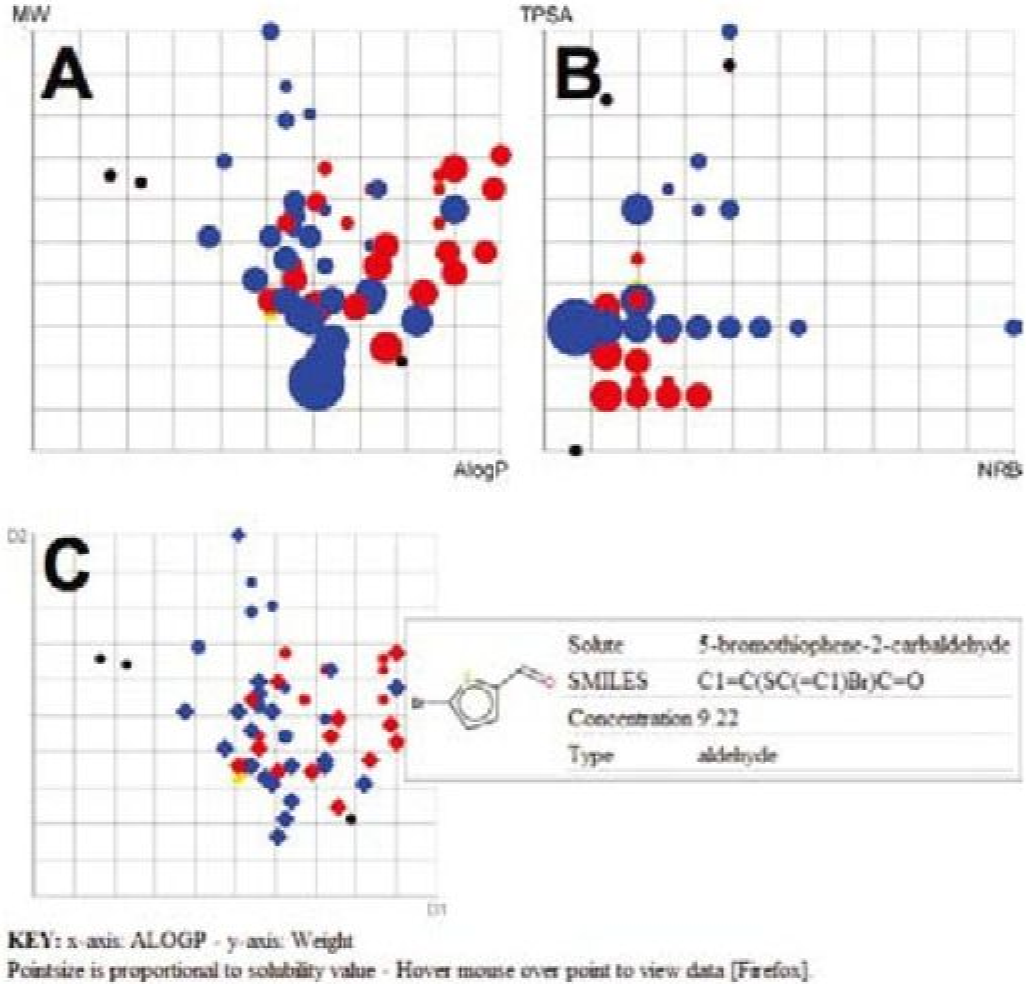
□□□□□B□Flickr□□□□□□□C□□□Gdoc□□□□□□□□□



□□ 54□□16-2□ □□□□□□□□□□□□□□□□□A□□□JavaScript□G□□Doc API□□□□□□□□□□□□□□□□B□□□□□□□□□□□□□□C□□□□□□□□□□□□□□□□□□□□□□□□http□//toposome.chemistry.drexel.edu/□rguha/jcsol/sol.html□□



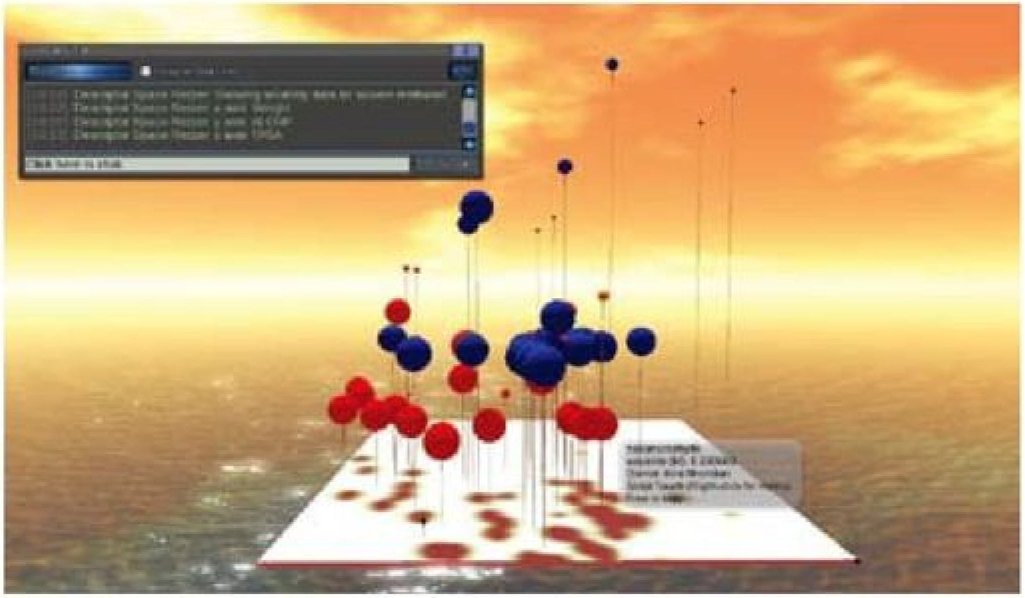
□□ 55□□16-3□ □□ChemSpider□□□□□□□□□□□□□□□□□□□□□


MW
A
AlogP
TPSA
B.
NRB
D2
C
D1
Solute 5-bromothiophene-2-carbaldehyde
SMILES C1=C(SC(=C1)Br)C=O
Concentration 9.22
Type aldehyde
KEY: x-axis: ALOGP - y-axis: Weight
Pointsize is proportional to solubility value - Hover mouse over point to view data [Firefox].

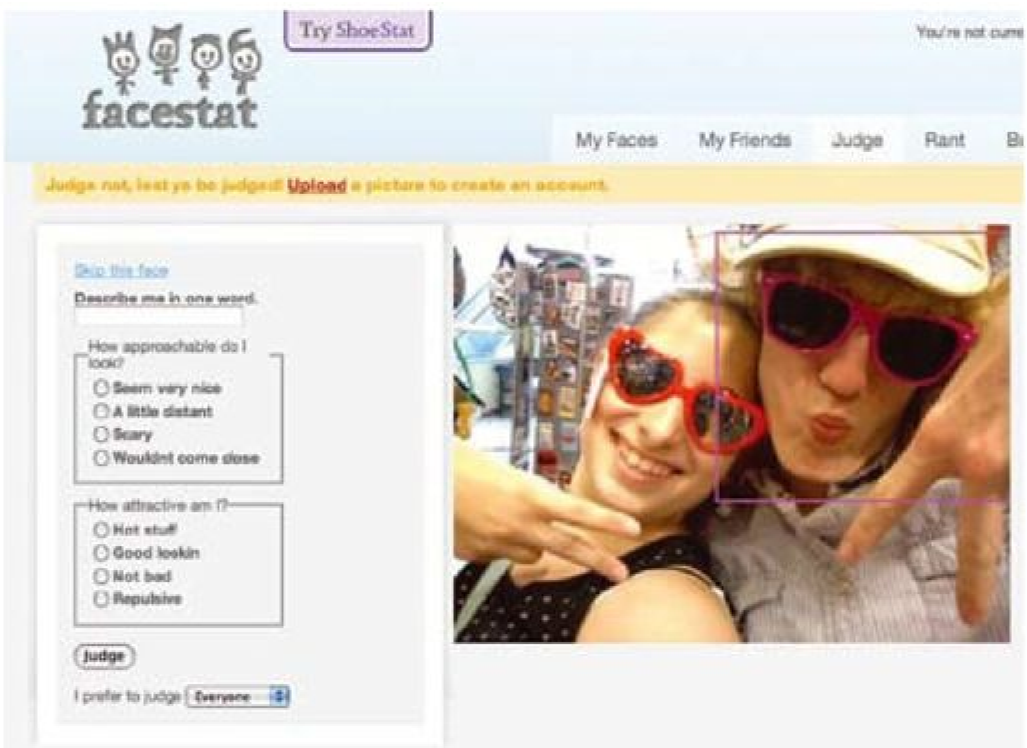
[illegible] 56[illegible]16-5[illegible] [illegible]A[illegible]B[illegible]

[illegible]

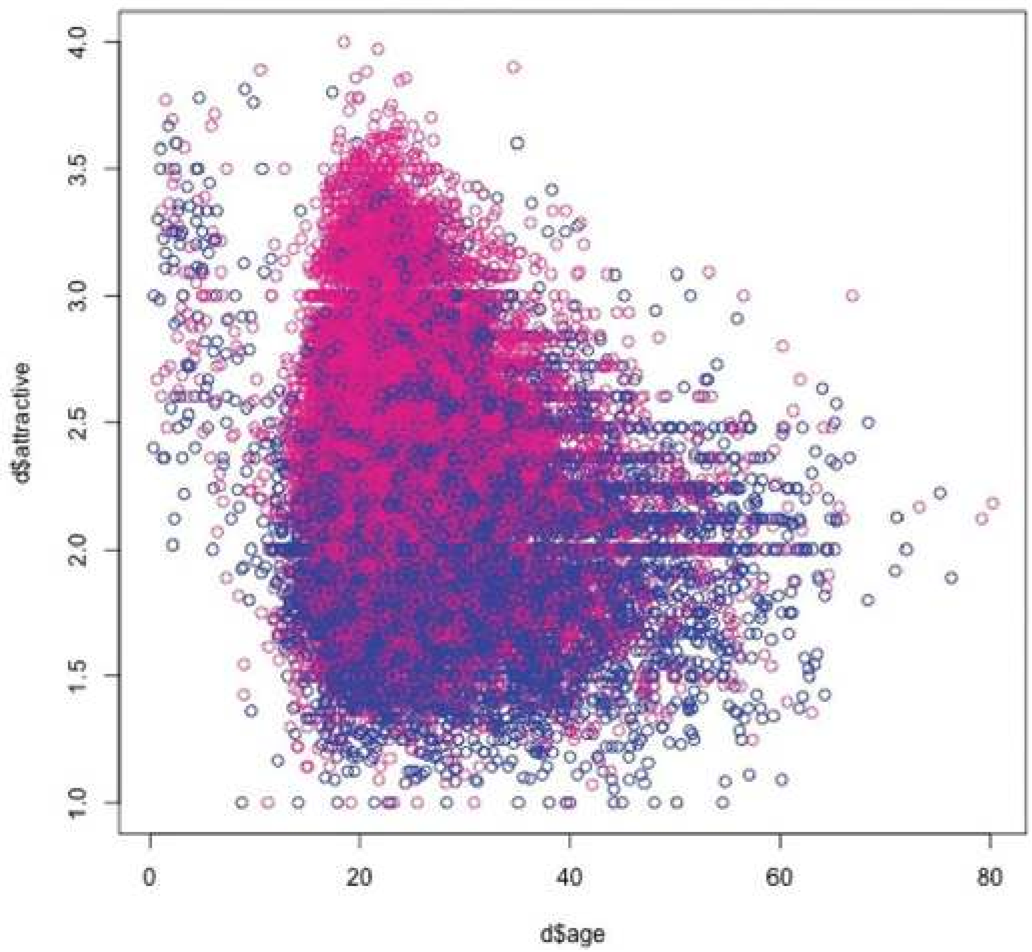
[illegible]C[illegible]



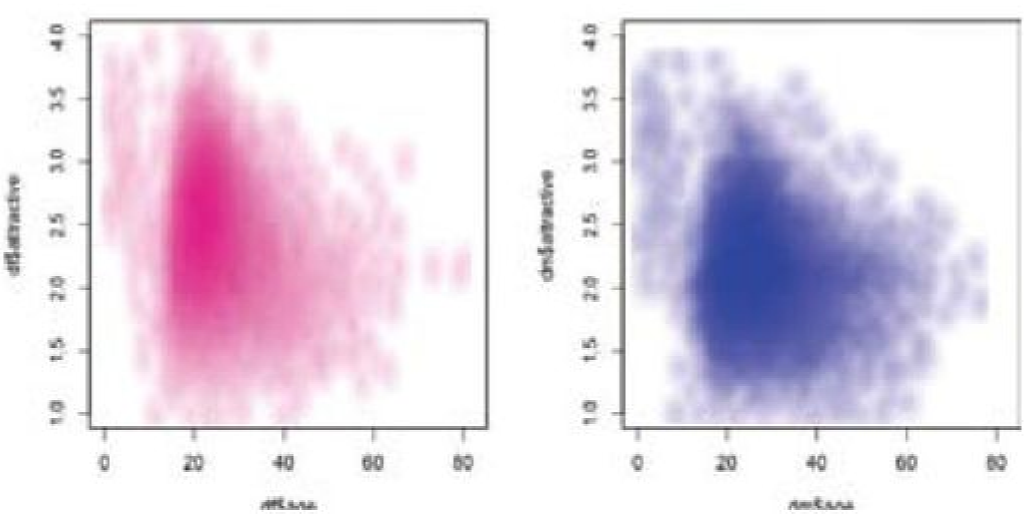
□□ 57□□16-6□ □□Second Life□□□16-5□□□□□□□□□□□□□□□□□□□□□□□□□□□□ http □ //slurl.com/secondlife/Drexel/165/178/24 □ □ □ □ Second Life□□Drexel□



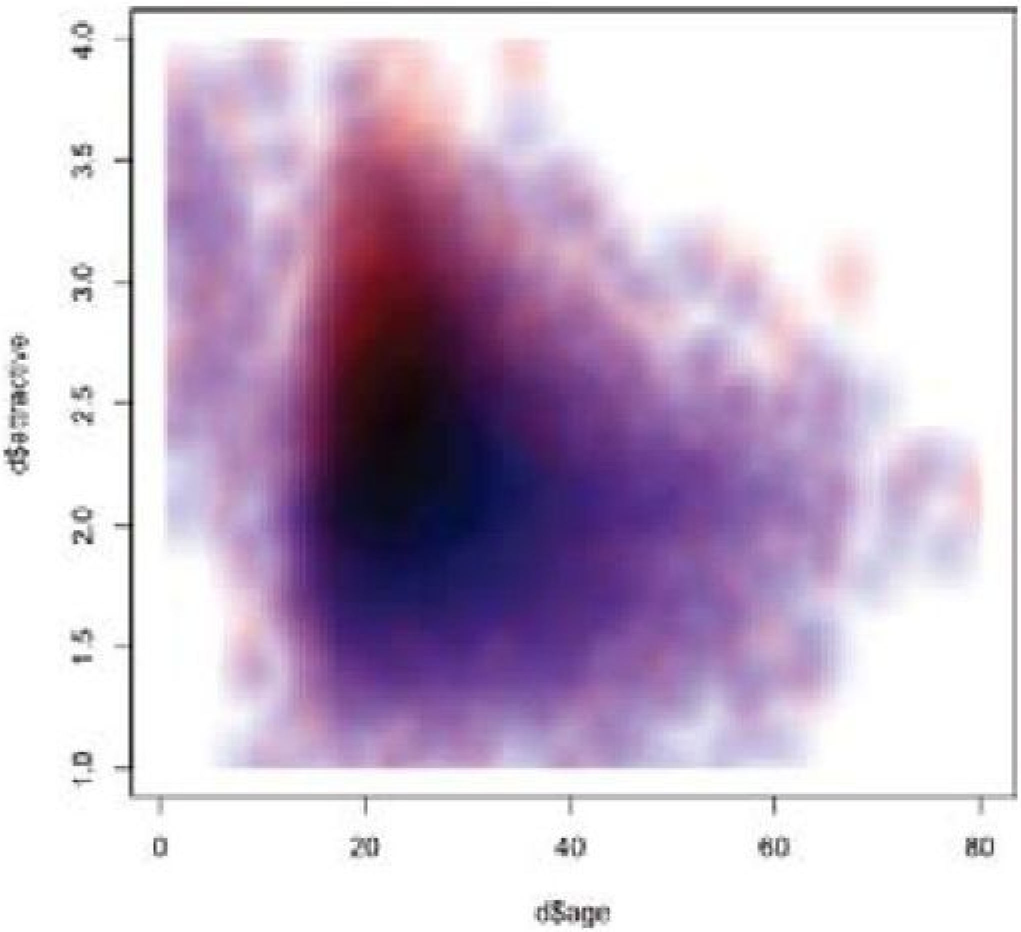
□□ 58□□17-1□ FaceStat□□□□□



□□ 59□□17-5□ □□□□□□□□□□□□□□□□□□□□□

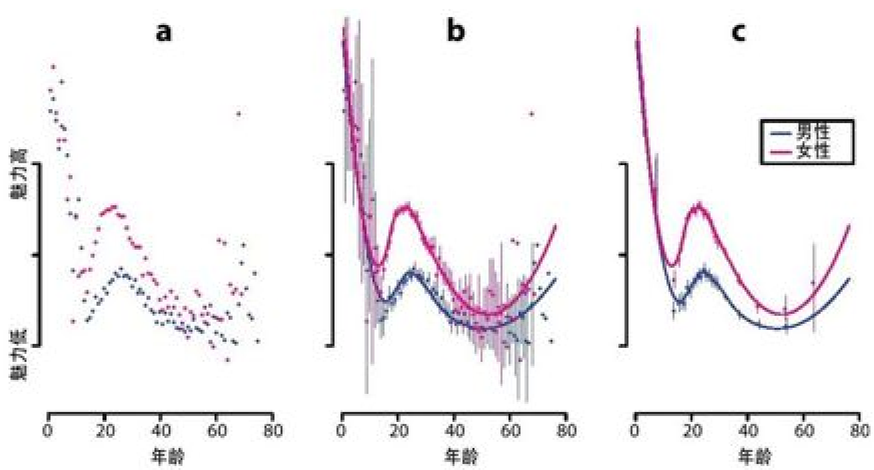


[illegible] 60[illegible]17-6[illegible] [illegible]



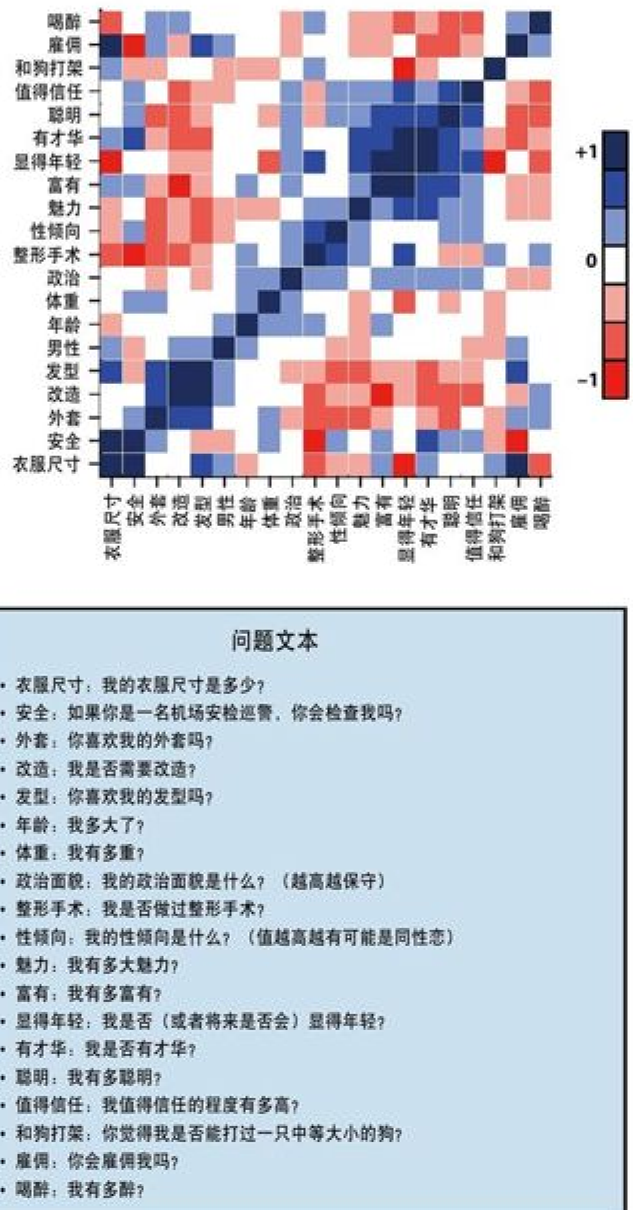
d$attractive
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0
20
40
60
80
d$age

[illegible] 61[illegible]17-7[illegible] [illegible]

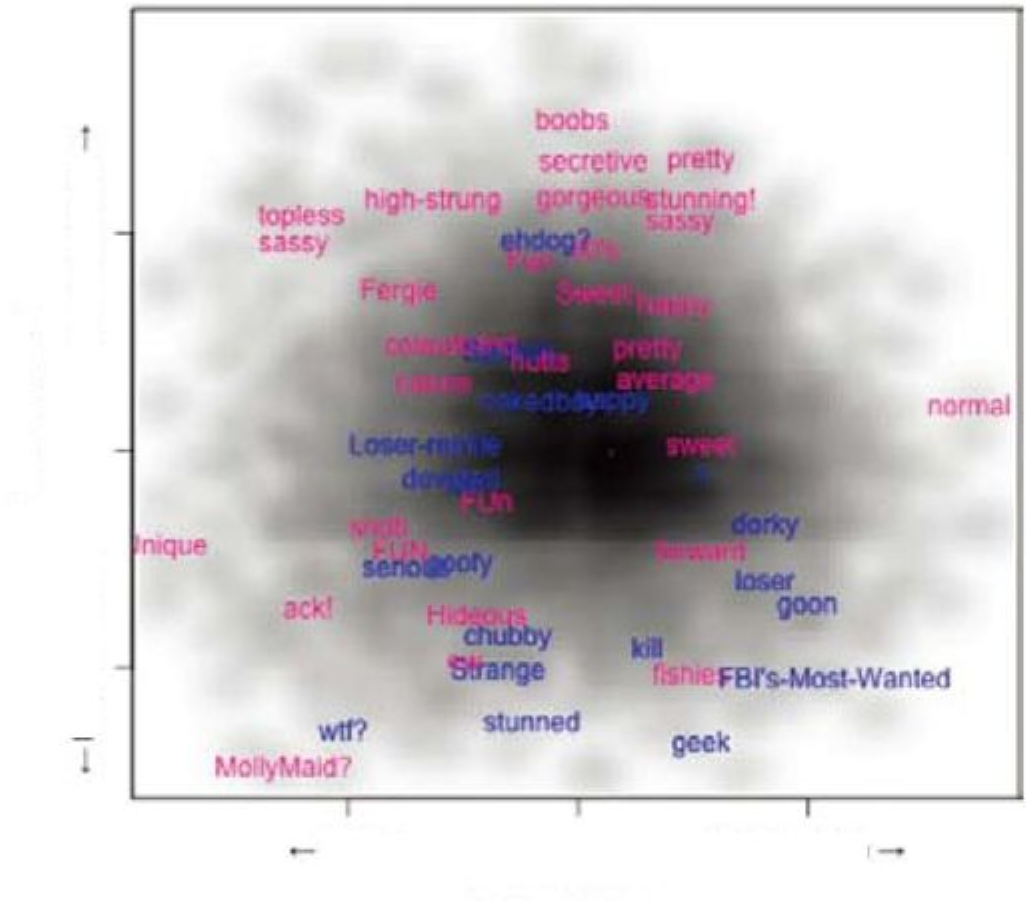


a
b
c
魅力高
魅力低
男性
女性
0
20
40
60
80
年齢

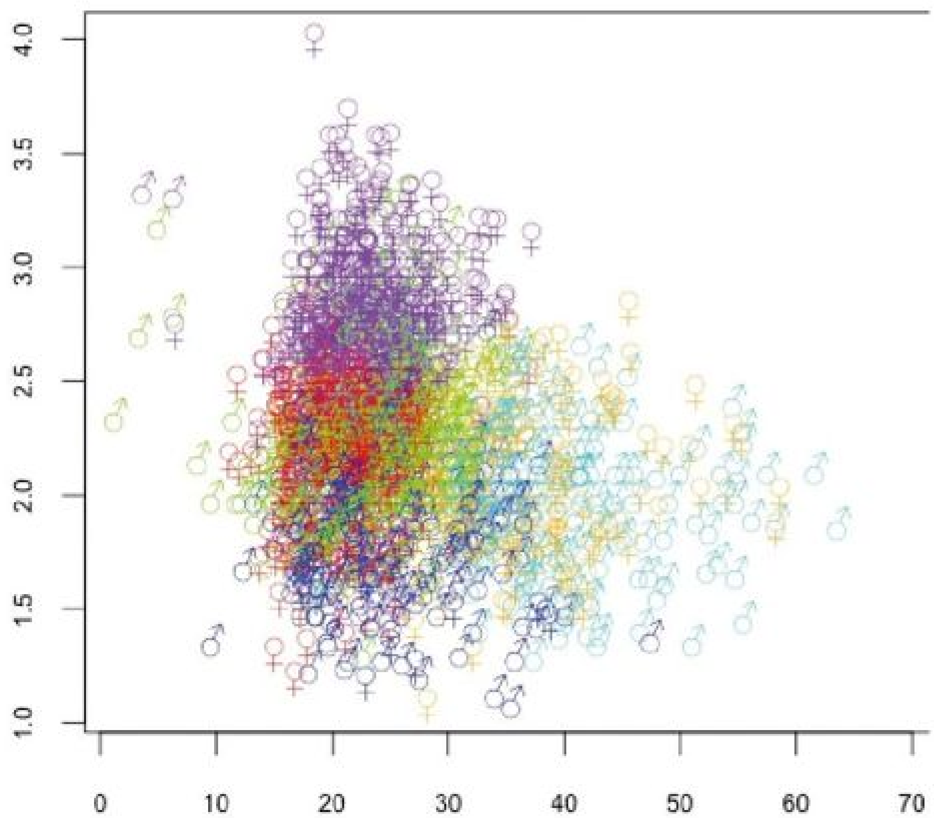
[illegible] 62[illegible]17-8[illegible] [illegible]a[illegible]b[illegible]95%[illegible]c[illegible]



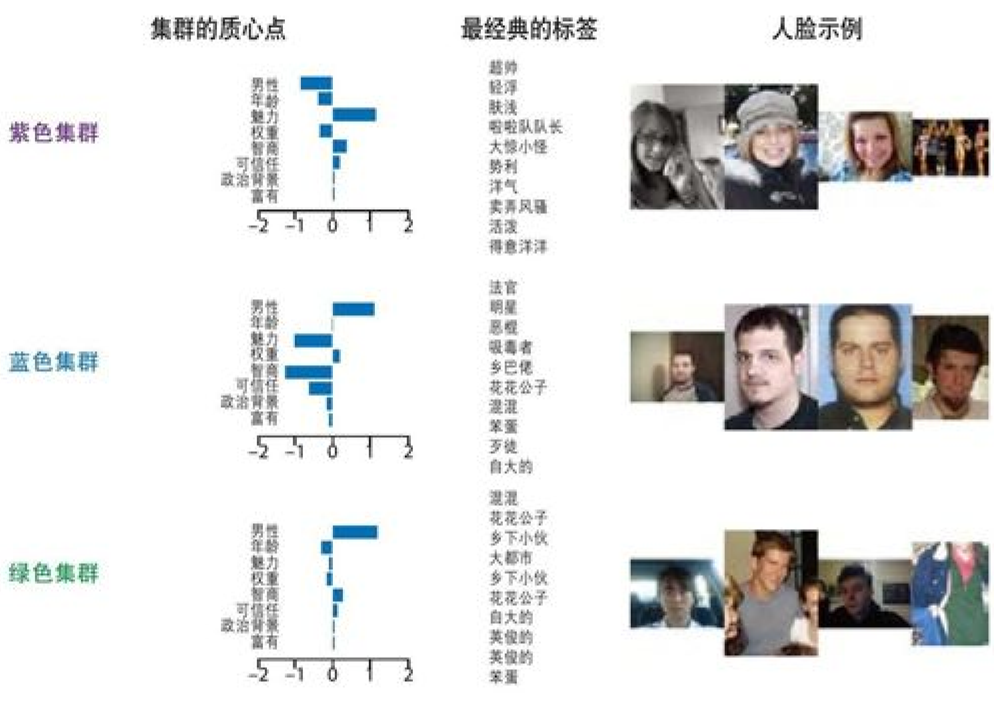
[illegible] 63[illegible]17-9[illegible] Pearson[illegible]


boobs
secretive
pretty
high-strung
gorgeous
stunning!
topless
sassy
sassy
ehdog?
Fergie
pretty
average
normal
Loser-ish
sweet
FUN
dorky
Unique
goofy
loser
goon
ack!
Hideous
chubby
kill
Strange
fishie
FBI's-Most-Wanted
stunned
wtf?
geek
MollyMaid?

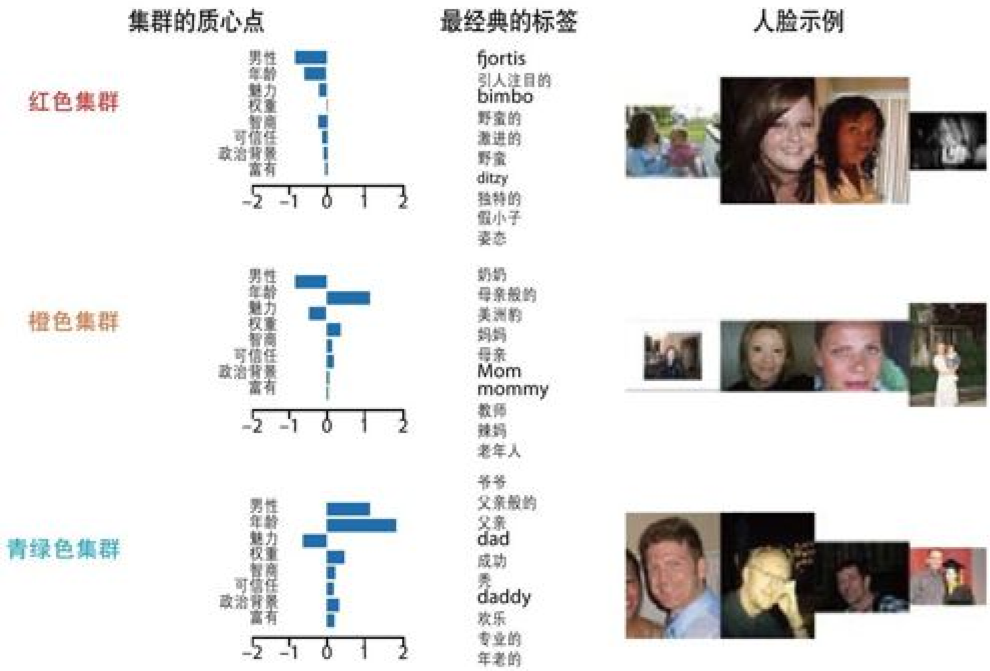
□□ 64□□17-12□ □□□□□□□□□□□□□□□□□□□


4.0
3.5
3.0
2.5
2.0
1.5
1.0
0
10
20
30
40
50
60
70

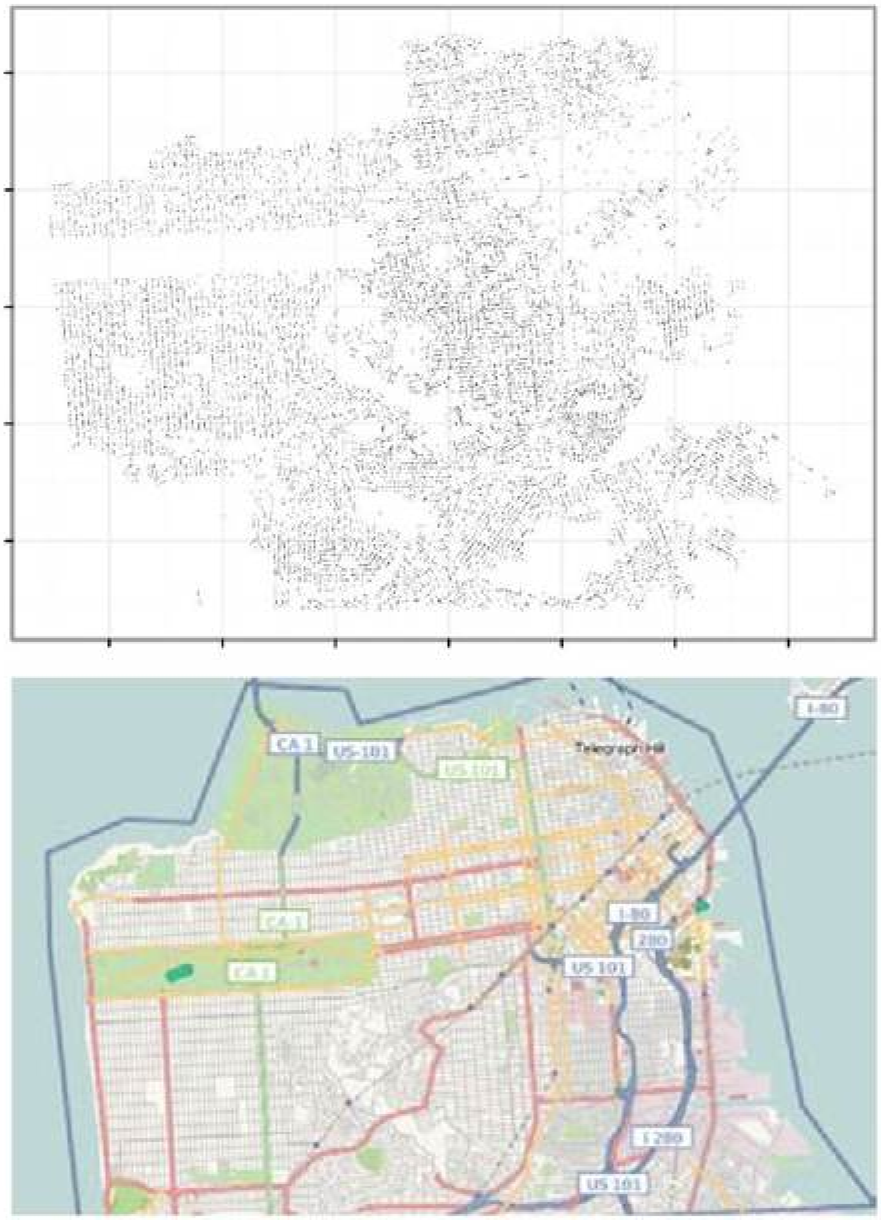
□□ 65□□17-13□ □□□□□□□□□□□□□□□□□□□□□2000□□□□□□


集群的质心点
最经典的标签
人脸示例
紫色集群
男性
年龄
魅力
权重
智商
可信任
政治背景
富有
−2 −1 0 1 2
超帅
轻浮
肤浅
啦啦队队长
大惊小怪
势利
洋气
卖弄风骚
活泼
得意洋洋
蓝色集群
男性
年龄
魅力
权重
智商
可信任
政治背景
富有
−2 −1 0 1 2
法官
明星
恶棍
吸毒者
乡巴佬
花花公子
混混
笨蛋
歹徒
自大的
绿色集群
男性
年龄
魅力
权重
智商
可信任
政治背景
富有
−2 −1 0 1 2
混混
花花公子
乡下小伙
大都市
乡下小伙
花花公子
自大的
英俊的
英俊的
笨蛋

□□ 66□□17-15□ □□□□□□□□□□□□



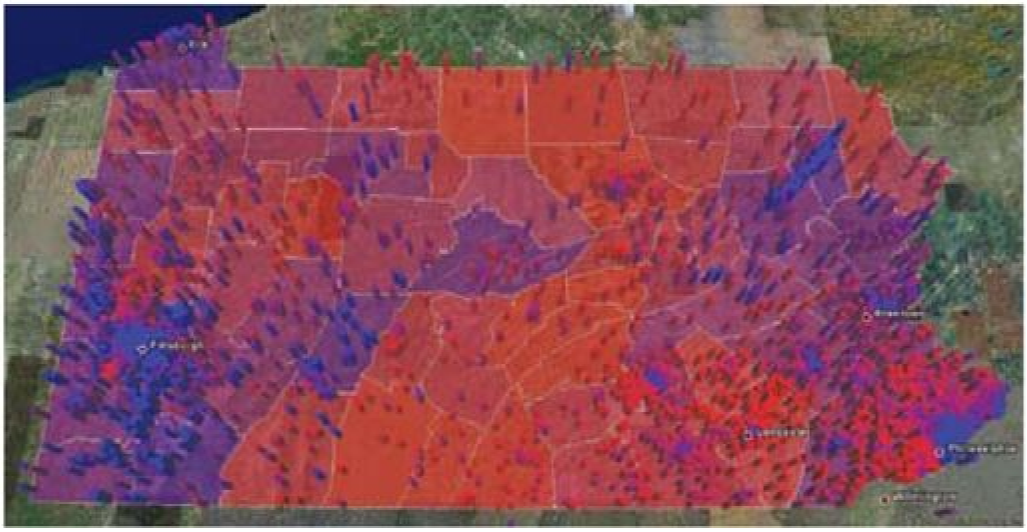
□□ 67□□17-16□ □□□□□□□□□□□□



[illegible] 68[illegible]18-13[illegible] [illegible]

[illegible]

http[illegible]//openstreetmap.com



[illegible] 69[illegible]19-5[illegible] [illegible]2004[illegible]

[illegible]John Kerry[illegible]

[illegible]George W.Bush[illegible]

[illegible]4000[illegible]1[illegible]

[illegible]1[illegible]

[illegible]——[illegible]——[illegible]

[illegible]



[illegible] 70[illegible]20-1[illegible] [illegible]